

(FILE 'HOME' ENTERED AT 17:04:08 ON 16 SEP 2002)

FILE 'MEDLINE, BIOSIS, CANCERLIT, LIFESCI, BIOTECHDS, CAPLUS' ENTERED AT  
17:04:34 ON 16 SEP 2002

L1           15 S DHELLIN?/AU  
L2           303 S AMIGORENA?/AU  
L3           249 S RAMEAU?/AU  
L4           1167 S CROUZET?/AU  
L5           1727 S L1 OR L2 OR L3 OR L4  
L6           67 S L5 AND (VESICLE# OR (MEMBRANE (3A) PREPAR?))  
L7           25 DUP REM L6 (42 DUPLICATES REMOVED)  
L8           6 S L5 AND (ANION(W) EXCHANGE?)  
L9           3 DUP REM L8 (3 DUPLICATES REMOVED)  
L10          26 S L7 OR L9  
L11          5 DUP REM L1 (10 DUPLICATES REMOVED)  
L12          315851 S (VESICLE# OR (MEMBRANE (3A) PREPAR?))  
L13          1888 S L12 AND (ANION(W) EXCHANGE?)  
L14          1236 S L12(S) (ANION(W) EXCHANGE?)  
L15          430 S L12(5A) (ANION(W) EXCHANGE?)  
L16          410 S L15 AND PY<2001  
L17          359 S L12(3A) (ANION(W) EXCHANGE?)  
L18          340 S L17 AND PY<2001  
L19          299 DUP REM L18 (41 DUPLICATES REMOVED)  
L20          337 S L13 AND (ANION(W) EXCHANGE#)/TI  
L21          102 S L20 AND L12/TI  
L22          100 S L21 AND PY<2001  
L23          70 DUP REM L22 (30 DUPLICATES REMOVED)  
L24          66557 S SYNAPTOSOM? OR ENDOSOM? OR EXOSOM?  
L25          64 S L24 AND (ANION(W) EXCHANGE#)  
L26          60 S L25 AND PY<2001  
L27          22 DUP REM L26 (38 DUPLICATES REMOVED)  
L28          2 S L13 AND DENDRI?  
L29          19 S L13 AND (LYMPHOCYT? OR (B(W)CELL#) OR (T(W)CELL#) OR APC OR  
(  
L30          18 S L29 AND PY<2001  
L31          10 DUP REM L30 (8 DUPLICATES REMOVED)  
L32          17 S L5 AND (ION(W) EXCHANG?)  
L33          16 S L32 NOT L8  
L34          11 DUP REM L33 (5 DUPLICATES REMOVED)  
L35          0 S L6 AND (DEAE OR QAE OR Q)  
L36          0 S L6 AND SEPHADEX  
L37          3033 S L12 AND (ION(W) EXCHANG?)  
L38          31 S L37 AND (DENDRI? OR LYMPHOCYT? OR (B(W)CELL#) OR (T(W)CELL#))  
L39          31 S L38 AND PY<2001  
L40          12 DUP REM L39 (19 DUPLICATES REMOVED)  
L41          244 S L24 AND (ION(W) EXCHANG?)  
L42          474 S L24 AND (DEAE OR QAE OR Q)  
L43          11 S L24(5A) (ION(W) EXCHANG?)  
L44          36 S L24(5A) (DEAE OR QAE OR Q)  
L45          47 S L43 OR L44  
L46          34 S L45 AND PY<2001  
L47          22 DUP REM L46 (12 DUPLICATES REMOVED)

# WEST Search History

DATE: Monday, September 16, 2002

<u>Set Name</u>	<u>Query</u>	<u>Hit Count</u>	<u>Set Name</u>
side by side			result set
<i>DB=JPAB,EPAB,DWPI; PLUR=NO; OP=ADJ</i>			
L17	vesicle\$1 and (((anion or ion) adj exchang\$2) or DEAE or QAE)	11	L17
L16	L13 and antigen\$2	8	L16
L15	L13 and (((anion or ion) adj exchang\$2) or DEAE or QAE)	0	L15
L14	L13 and (((anion or ion) adj exchang\$2) or DEAE or QAE or Q)	6	L14
L13	exosom\$2 or synaptosom\$2 or dendrisom\$2	81	L13
<i>DB=USPT; PLUR=NO; OP=ADJ</i>			
L12	L11 and @ad<20000119	44	L12
L11	L5 with (ion adj exchang\$3)	44	L11
L10	L9 or l8	44	L10
L9	L7 and @prad<20000119	3	L9
L8	L7 and @ad<20000119	44	L8
L7	L5 same (anion adj exchang\$3)	45	L7
L6	L5 with (anion adj exchang\$3)	9	L6
L5	vesicle\$1 or endosom\$2 or synaptosom\$2 or exosom\$2	11442	L5
<i>DB=USPT,PGPB; PLUR=NO; OP=ADJ</i>			
L4	vesicle\$1 or endosom\$2 or synaptosom\$2 or exosom\$2	13742	L4
L3	L1 and (anion adj exchang\$3)	4	L3
L2	L1 and vesicle\$1	2	L2
L1	dhellin\$[in] or amigorena\$[in] or rameau\$[in] or crouzet\$[in]	55	L1

END OF SEARCH HISTORY

L10 ANSWER 6 OF 80 MEDLINE DUPLICATE 6  
ACCESSION NUMBER: 2000132867 MEDLINE  
DOCUMENT NUMBER: 20132867 PubMed ID: 10666310  
TITLE: Evidence for requirement of NADPH-cytochrome P450  
oxidoreductase in the microsomal NADPH-sterol  
Delta7-reductase system.  
AUTHOR: Nishino H; Ishibashi T  
CORPORATE SOURCE: Department of Biochemistry, Hokkaido University School of  
Medicine, Sapporo, 060-8638, Japan.  
SOURCE: ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS, (2000 Feb 15) 374  
(2) 293-8.  
Journal code: 0372430. ISSN: 0003-9861.  
PUB. COUNTRY: United States  
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)  
LANGUAGE: English  
FILE SEGMENT: Priority Journals  
ENTRY MONTH: 200003  
ENTRY DATE: Entered STN: 20000327  
Last Updated on STN: 20000327  
Entered Medline: 20000315  
AB Rabbit antibodies raised against the hydrophilic part of  
**microsomal NADPH-cytochrome** P450 oxidoreductase (denoted  
fpT) demonstrated a marked ability to inhibit NADPH-sterol  
Delta7-reductase activity. In addition, trypsin and proteinase K  
treatment  
of **microsomes** removed almost all **microsomal** electron  
transfer constituents from the **microsomes**, but the  
Delta7-reductase activity could be reconstituted by adding  
detergent-solubilized NADPH-cytochrome P450 oxidoreductase  
(denoted OR). Furthermore, after solubilization from **microsomes**,  
the Delta7-reductase activity could be reconstituted with OR in a  
DEAE-cellulose column chromatography eluate fraction, which contained  
little OR activity. In the **microsomal** system, carbon monoxide,  
ketoconazole, and miconazole, specific inhibitors of **cytochrome**  
P450, had no effect on Delta7-reductase activity. These results provide  
the first evidence of an essential requirement of OR, which is distinct  
from **cytochrome** P450, in the NADPH-sterol Delta7-reductase  
system. EDTA, o-phenanthroline and KCN markedly lowered Delta7-reductase  
activity in a dose-dependent manner. Among metal ions tested, only ferric  
ion restored the reductase activity in the EDTA-treated **microsomes**  
. These results suggest that NADPH-sterol Delta7-reductase is  
membrane-bound iron-dependent protein embedded in the **microsomal**  
**lipid bilayer.**  
Copyright 2000 Academic Press.

L5 ANSWER 2 OF 15 MEDLINE DUPLICATE 2  
ACCESSION NUMBER: 87109259 MEDLINE  
DOCUMENT NUMBER: 87109259 PubMed ID: 3100531  
TITLE: Immunochemical and kinetic evidence for two different prostaglandin H-prostaglandin E isomerasases in sheep vesicular gland microsomes.  
AUTHOR: Tanaka Y; Ward S L; Smith W L  
CONTRACT NUMBER: AM22042 (NIADDK)  
HL07404 (NHLBI)  
SOURCE: JOURNAL OF BIOLOGICAL CHEMISTRY, (1987 Jan 25) 262 (3)  
1374-81.  
Journal code: 2985121R. ISSN: 0021-9258.  
PUB. COUNTRY: United States  
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)  
LANGUAGE: English  
FILE SEGMENT: Priority Journals  
ENTRY MONTH: 198703  
ENTRY DATE: Entered STN: 19900303  
Last Updated on STN: 19970203  
Entered Medline: 19870304  
AB Splenic lymphocytes from mice immunized with a partially purified prostaglandin (PG) H-PGE isomerase from sheep vesicular glands were fused with SP2/0-Ag14 myeloma cells. Two spleen cell-myeloma hybrids (hei-7 and hei-26) were selected and cloned. The mouse antibodies secreted by the two hybrids, IgG1 (hei-7) and IgG1 (hei-26), caused immunoprecipitation of a maximum of 45 and 22%, respectively, of the solubilized PGH-PGE isomerase activity of sheep vesicular gland; immunoprecipitation of activity by the two antibodies was additive. The antigens reactive with IgG1 (hei-7) and IgG1 (hei-26) were identified as proteins with Mr = 17,500 and 180,000, respectively, by Western transfer blotting or sodium dodecyl sulfate-polyacrylamide gel electrophoresis of immunoprecipitated <sup>125</sup>I-labeled microsomes. The PGH-PGE isomerase activities precipitated by IgG1 (hei-7) and IgG1 (hei-26) exhibited different kinetic properties with respect to time course, Km for PGH<sub>2</sub>, and concentration dependence for GSH. No significant GSH-S-transferase activity was present in these immunoprecipitates. These data indicate that there are at least two different proteins in sheep vesicular gland microsomes capable of catalyzing GSH-dependent PGH-PGE isomerase reactions. IgG1 (hei-7), but not IgG1 (hei-26), caused coprecipitation of PGH synthase and PGH-PGE isomerase activities when incubated with intact right-side-out vesicular gland microsomes. Thus, the epitope for IgG1 (hei-7) is located on the **cytoplasmic** surface of those **microsomal** spheres which contain PGH synthase. This latter finding suggests that the isomerase reactive with IgG1 (hei-7) is involved in PGE synthesis in sheep vesicular glands.

L5 ANSWER 3 OF 15 MEDLINE DUPLICATE 3

L5 ANSWER 15 OF 15 CAPLUS COPYRIGHT 2002 ACS  
ACCESSION NUMBER: 1952:27303 CAPLUS  
DOCUMENT NUMBER: 46:27303  
ORIGINAL REFERENCE NO.: 46:4649f-h  
TITLE: The function of cytoplasmic granules and their significance in carcinogenic degeneration  
AUTHOR(S): Seeger, P. G.  
SOURCE: Z. Krebsforsch. (1950), 57, 113-20  
DOCUMENT TYPE: Journal  
LANGUAGE: Unavailable  
AB Ideas on mechanisms of carcinogenesis are presented. Intracellular biocatalysts such as vitamins, hormones, and enzymes are located in the **cytoplasmic** granules (mitochondria and **microsomes**); each cell **contains** a certain ratio of male and female hormone. The reaction of mutated or foreign microsomes with estrogen present in granules causes a multiplication of the cellular microsomes and an increase in ribonucleoprotein and protein synthesis. This would lead successively to nuclear multiplication, pathol. mitoses, and unrestricted cell growth. The cell might react by an increase in cholesterol, regulator of enzymic processes.

L11 ANSWER 20 OF 23 MEDLINE DUPLICATE 5  
ACCESSION NUMBER: 80157370 MEDLINE  
DOCUMENT NUMBER: 80157370 PubMed ID: 6153999  
TITLE: A new cellulose-based microcarrier culturing system.  
AUTHOR: Reuveny S; Bino T; Rosenberg H; Mizrahi A  
SOURCE: DEVELOPMENTS IN BIOLOGICAL STANDARDIZATION, (1980) 46  
137-45.  
Journal code: 0427140. ISSN: 0301-5149.  
PUB. COUNTRY: Switzerland  
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)  
LANGUAGE: English  
FILE SEGMENT: Priority Journals  
ENTRY MONTH: 198006  
ENTRY DATE: Entered STN: 19900315  
Last Updated on STN: 19970203  
Entered Medline: 19800625  
AB A search for new substrates to be used as microcarriers for culturing mammalian cells was carried out. Commercially available microgranular anion exchange DEAE-cellulose ( DE-52 of Whatman) were investigated as microcarriers for anchorage-dependent-cells. Cells from CCL-1 mouse cell line were grown on the investigated microcarriers. Mouse interferon was successfully produced after induction with Sendai virus. Interferon yield per cell was similar to that obtained in monolayer culture.

L5 ANSWER 13 OF 14 MEDLINE  
ACCESSION NUMBER: 81197352 MEDLINE  
DOCUMENT NUMBER: 81197352 PubMed ID: 6112703  
TITLE: Purification and characterization of synaptic vesicles  
from  
the electric organ of *Torpedo ocellata*.  
AUTHOR: Michaelson D M; Ophir I  
SOURCE: MONOGRAPHS IN NEURAL SCIENCES, (1980) 7 19-29.  
Journal code: 0357002. ISSN: 0300-5186.  
PUB. COUNTRY: Switzerland  
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)  
LANGUAGE: English  
FILE SEGMENT: Priority Journals  
ENTRY MONTH: 198107  
ENTRY DATE: Entered STN: 19900316  
Last Updated on STN: 19900316  
Entered Medline: 19810723

AB Highly **purified** synaptic **vesicles** have been isolated from the electric organ of *Torpedo ocellata* by a rapid procedure which enables the concurrent isolation of synaptic vesicles and of intact presynaptic nerve endings (**synaptosomes**). The **purification** procedure consists of homogenization of fresh electric tissue in iso-osmotic glycine in the presence of EGTA, differential and density gradient centrifugation, and **gel permeation** on a glass beads column of 2500 a pore size. The purity of the vesicles was evaluated both biochemically and morphologically. The vesicles contain acetylcholine (ACh) and ATP in a ratio of 3:1 and at specific concentrations of 2,100nmol ACh/mg protein and 1,010nmol ACh/mg phospholipid. They are associated with Ca<sup>+2</sup>/Mg<sup>+2</sup> ATPase activity and are devoid of the ouabain sensitive Na<sup>+</sup>/K<sup>+</sup> ATPase.

The relatively high yields as well as the short preparation time (about 9h for the vesicles and 4h for the synaptosomes) enables the employment of large samples of the isolated material on the day of preparation.

L74 ANSWER 16 OF 16 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC.  
ACCESSION NUMBER: 1987:370888 BIOSIS  
DOCUMENT NUMBER: BR33:61363  
TITLE: COMBINATION CHEMOIMMUNOTHERAPY AND IMMUNOTHERAPY FOR  
**METASTATIC** DISEASE UTILIZATION OF RH TNF RM  
IFN-ALPHA RM CSF GM AND RH IL-2.  
AUTHOR(S): TALMADGE J E; TRIBBLE H; PHILLIPS H; SCHNEIDER M;  
PENNINGTON R; LENZ B  
CORPORATE SOURCE: PRECLINICAL SCREENING LAB., PRI, NCI-FCRF, P.O. BOX B,  
FREDERICK, MD. 21701.  
SOURCE: SEVENTY-EIGHTH ANNUAL MEETING OF THE AMERICAN ASSOCIATION  
FOR CANCER RESEARCH, ATLANTA, GEORGIA, USA, MAY 20-23,  
1987. PROC AM ASSOC CANCER RES ANNU MEET, (1987) 28 (0),  
399.  
CODEN: PAMREA.  
DOCUMENT TYPE: Conference  
FILE SEGMENT: BR; OLD  
LANGUAGE: English

**PCT**
**WORLD INTELLECTUAL PROPERTY ORGANIZATION**  
 International Bureau

**INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)**

(51) International Patent Classification <sup>6</sup> :		A2	(11) International Publication Number: <b>WO 95/32292</b>
C12N 15/40, G01N 33/576, C07K 14/18, C12P 21/08, C07K 16/10			(43) International Publication Date: 30 November 1995 (30.11.95)
(21) International Application Number: PCT/US95/06266		(71) Applicant ( <i>for all designated States except US</i> ): GENELABS TECHNOLOGIES, INC. [US/US]; 505 Penobscot Drive, Redwood City, CA 94063 (US).	
(22) International Filing Date: 17 May 1995 (17.05.95)		(72) Inventors; and	
(30) Priority Data:		(75) Inventors/Applicant's ( <i>for US only</i> ): FRY, Kirk E. [US/US]; 2604 Ross Road, Palo Alto, CA 94303 (US). KIM, Jungsu, P. [US/US]; 1844 Guinda Street, Palo Alto, CA 94306 (US). MURPHY, Frederick A. [US/US]; 27324 Golf View Circle, El Macero, CA 95618 (US). LINNEN, Jeffrey M. [US/US]; 1017 Catamaran Street #1, Foster City, CA 94404 (US).	
08/246,985 20 May 1994 (20.05.94) US 08/285,561 3 August 1994 (03.08.94) US 08/329,729 26 October 1994 (26.10.94) US 08/344,271 23 November 1994 (23.11.94) US 08/357,509 16 December 1994 (16.12.94) US 08/389,886 15 February 1995 (15.02.95) US		(74) Agent: FABIAN, Gary R.; Dehlinger & Associates, P.O. Box 60850, Palo Alto, CA 94306-0850 (US).	
(60) Parent Applications or Grants		(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).	
(63) Related by Continuation		Published <i>Without international search report and to be republished upon receipt of that report.</i>	
US	08/246,985 (CIP)		
Filed on	20 May 1994 (20.05.94)		
US	08/285,561 (CIP)		
Filed on	3 August 1994 (03.08.94)		
US	08/329,729 (CIP)		
Filed on	26 October 1994 (26.10.94)		
US	08/344,271 (CIP)		
Filed on	23 November 1994 (23.11.94)		
US	08/357,509 (CIP)		
Filed on	16 December 1994 (16.12.94)		
US	08/389,886 (CIP)		
Filed on	15 February 1995 (15.02.95)		
(54) Title: DETECTION OF VIRAL ANTIGENS CODED BY REVERSE-READING FRAMES			
(57) Abstract			
<p>The present invention describes a novel method to determine whether a test subject is infected with a selected virus, where the virus has an RNA genome. The method includes the identification of polypeptide antigens coded by reverse open reading frames, that is, reading frames coded in the opposite direction to the major known viral reading frames. Further, the invention includes the reverse frame polypeptide antigens, methods of identifying and producing such polypeptide antigens, and antibodies that are specifically immunoreactive with said polypeptide antigens. These polypeptide antigens and antibodies are useful in diagnostic and therapeutic applications.</p>			

L42 ANSWER 8 OF 51 PCTFULL COPYRIGHT 2003 Univentio  
ACCESSION NUMBER: 1995032292 PCTFULL ED 20020514  
TITLE (ENGLISH): DETECTION OF VIRAL ANTIGENS CODED BY REVERSE-READING  
FRAMES  
TITLE (FRENCH): DETECTION D'ANTIGENES VIRAUX CODES PAR DES CADRES DE  
LECTURE INVERSE  
INVENTOR(S): FRY, Kirk, E.;  
KIM, Jungsuh, P.;  
MURPHY, Frederick, A.;  
LINNEN, Jeffrey, M.  
PATENT ASSIGNEE(S): GENELABS TECHNOLOGIES, INC.;  
FRY, Kirk, E.;  
KIM, Jungsuh, P.;  
MURPHY, Frederick, A.;  
LINNEN, Jeffrey, M.  
LANGUAGE OF PUBL.: English  
DOCUMENT TYPE: Patent  
PATENT INFORMATION:

NUMBER	KIND	DATE
WO 9532292	A2	19951130

## DESIGNATED STATES

W:

AM AT AU BB BG BR BY CA CH CN CZ DE DK EE ES FI GB GE  
HU JP KE KG KP KR KZ LK LR LT LU LV MD MG MN MW MX NO  
NZ PL PT RO RU SD SE SI SK TJ TT UA US US US US US  
UZ VN KE MW SD SZ UG AT BE CH DE DK ES FR GB GR IE IT  
LU MC NL PT SE BF BJ CF CG CI CM GA GN ML MR NE SN TD  
TG

## APPLICATION INFO.:

WO 1995-US6266 A 19950517  
US 1994-8/246,985 19940520  
US 1994-8/285,561 19940803  
US 1994-8/329,729 19941026  
US 1994-8/344,271 19941123  
US 1994-8/357,509 19941216  
US 1995-8/389,886 19950215

AI WO 1995-US6266 A 19950517

DETD . . . lysed

cells or culture media. Purification can be carried out by methods known in the art including salt fractionation, ion exchange chromatography, and **affinity chromatography**,

Immunoaffinity chromatography can be employed using antibodies generated based on the HGV antigens identified by the methods of the present invention, HGV polypeptide antigens. . . multiple, tandem epitopes can be constructed that will produce mosaic proteins using standard recombinant DNA technology using polypeptide expression vector/host system described above. Further, **multiple antigen peptides** can be synthesized

chemically by methods described previously (Tam, J.P., 1988; Briand et al.). For example, a small immuno- logically inert core. . . used to anchor multiple copies of the same or different synthetic peptides (typically 6-15 residues long) representing epitopes of interest, Mosaic proteins or **multiple antigen peptide**

**antigens** give higher sensitivity and specificity in immunoassays due to the signal amplification resulting from distribution of multiple epitopes.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Makawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Togo
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States
FI	Finland	MN	Mongolia	UZ	Uzbekistan

**DETECTION OF VIRAL ANTIGENS  
CODED BY REVERSE-READING FRAMES**

**5 FIELD OF INVENTION**

This invention relates to a novel method to determine whether a subject is infected with a virus. The method includes the use of antigens coded by reverse open reading frames, that is, reading frames coded in the opposite direction to the major known viral reading frames. Also included in the invention are the reverse frame antigens, methods of identifying and producing such antigens, and antibodies that are specifically immunoreactive with said antigens. The invention also relates to diagnostic and therapeutic methods involving these antigens and antibodies.

**REFERENCES**

- Abstracts, The 1992 San Diego Conf.: Genetic Recognition, *Clin. Chem.* 39(4):705 (1993)
- Alter, H.J., Abstracts of Int. Symp. on Viral Hepatitis and Liver Dis., p. 47 (1993).
- Altschul, S., et al., *J. Mol. Biol.* 215:403-10 (1990).
- Ascadi, G., et al., *Nature* 352:815 (1991).
- Ausubel, F.M., et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley and Sons, Inc., Media PA.
- Barany, F., *PCR Methods Appl.* 1:5 (1991).
- Baron, S., et al., *JAMA* 266:1375 (1991).
- Beames, et al., *Biotechniques* 11:378 (1991).
- Blackburn, G.F., et al., *Clin. Chem.* 37:1534-1539 (1991).
- Bradley, D.W., et al., *J. Infec. Dis.*, 148:2 (1983).
- Bradley, D.W., et al., *J Gen. Virol.*, 69:1 (1988).
- Bradley, D.W. et al., *Proc. Nat. Acad. Sci., USA*, 84:6277 (1987).
- Briand, J.-P., et al., *J. Immunol. Meth.* 156:255 (1992).
- Cahill, P., et al., *Clin. Chem.* 37:1482 (1991).

- Cathal, G., et al., *DNA* 2:329-335 (1983).
- Chambers, T.J., et al., *Ann. Rev. Microbiol.* 44:649 (1990).
- Chambers, T.J., et al., *PNAS* 87:8898 (1990).
- 5 Chomczynski et al, *Anal. Biochem.* 162:159 (1987).
- Christian, R.B., et al., *J. Mol. Biol.* 227:771 (1992).
- Crea, R., U.S. Patent No. 4,888,286, issued December 19, 1989.
- 10 DeGraaf, M.E., et al., *Gene* 128:13 (1993).
- DiBisceglie, A.M., et al., *Hepatology* 16:649 (1992).
- DiBisceglie, A.M., et al., *NEJM* 321:1506 (1989).
- DiCesare, J., et al., *Biotechniques* 15:152-157 (1993).
- 15 Dienstag, J.L., et al, *Sem Liver Disease* 6:67 (1986).
- Eaton, M. A. W., et al., U.S. Patent No. 4,719,180, issued Jan. 12, 1988.
- Egholm, et al., *Nature* 365:566 (1993).
- EPO patent application 88310922.5, filed 11/18/88.
- 20 Farci, P., et al., *NEJM* 330:88 (1994).
- Felgner and Rhodes, *Nature* 349:251 (1991).
- Folgori, A., et al., *EMBO J.* 13:2236 (1994).
- Fracncki, R.I.B., et al., *Arch. Virol. Suppl* 12:223 (1991).
- 25 Frank, R., and Doring, R., *Tetrahedron* 44:6031-6040 (1988).
- Geysen, M., et al., *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984).
- Gingeras, T.R., et al., *Ann. Biol. Clin.* 48:498 (1990).
- 30 Gingeras, T.R., et al., *J. Inf. Dis.* 164:1066 (1991).
- Goeddel, D.V., *Methods in Enzymology* 185 (1990).
- Grakoui, A., et al., *J. Virol.* 67:2832 (1993).
- Guatelli, J.C., et al., *Proc. Natl. Acad. Sci. USA* 87:1874 (1990).
- 35 Gubler, U., et al, *Gene*, 25:263 (1983).

- Guthrie, C., and G.R. Fink, *Methods in Enzymology* 194 (1991).
- Gutterman, J.U., *PNAS* 91:1198 (1994).
- Harlow, E., et al., ANTIBODIES: A LABORATORY MANUAL, Cold Spring Harbor Laboratory Press (1988).
- Hieter, P.A., et al., *Cell* 22:197-207 (1980).
- Hijikata, M., et al., *PNAS* 88:5547 (1991).
- Hochuli, E., in GENETIC ENGINEERING, PRINCIPALS AND PRACTICE, VOL. 12 (J. Stelow Ed.) Plenum, NY, pp. 87-98 (1990).
- Holodniy, M., et al., *Biotechniques* 12:36 (1992).
- Hopp, T.P., et al., *Proc. Natl. Acad. Sci. USA* 78:3824-3828 (1981).
- Horn, T., and Urdea, M.S., *Nuc. Acids. Res.* 17:6959 (1989).
- Houghten, R.A., *Proc. Natl. Acad. Sci. USA* 82:5131 (1985).
- Hudson, D., *J. Org. Chem.* 53:617 (1988).
- Irwin, M.J., et al., *J. Virol.* 58:5036 (1994).
- Jacob, J.R., et al., in THE MOLECULAR BIOLOGY OF HCV, Section 4, pages 387-392 (1991).
- Jacob, J.R., et al., *Hepatology* 10:921-927 (1989).
- Jacob, J.R., et al., *J. Infect. Dis.* 161:1121-1127 (1990).
- Kakumu, S., et al., *Gastroenterol.* 105:507 (1993).
- Katz, E.D., and Dong, M., *Biotechniques* 8:546 (1990).
- Kawasaki, E.S., et al., in PCR TECHNOLOGY: PRINCIPLES AND APPLICATIONS OF DNA AMPLIFICATION (H.A. Erlich, ed.) Stockton Press (1989).
- Koonin, E.V., and Dolja, V.V., *Critical Reviews in Biochem. & Mol. Biol.* 28:375-430 (1993).
- Krausslich, H.G., et al., VIRAL PROTEINASES AS TARGETS FOR CHEMOTHERAPY (Cold Spring Harbor Press, Plainville, NY) (1989).
- Kumar, R., et al., *AIDS Res. Human Retroviruses* 5(3):345-354 (1989).
- Larder, B.A., and Kemp, S.D., *Science* 246:1155 (1989).

- 4
- WO 95/32292  
PCT/US95/06266
- 25:174-182 (1989).  
Lanford, R.E., et al., *In Vitro Cell. Dev. Biol.*  
Lomell, H., et al., *Clin. Chem.* 48:492 (1990).  
Maniatis, T., et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory (1982).  
Marshall, W.S., and Garthwaite, M.H., *Science* 259:1564  
Messing, J., *Methods in Enzymol.* 201:20 (1983).  
Michele, et al., *International Symposium on Viral*  
Miller, J. H., *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory (1972).  
Mortisssey, D.V., et al., *Anal. Biochem.* 181:345  
Moss, B., et al., *CURRENT Protocols in Molecular Biology* (1989).  
5 MANUAL, Cold Spring Harbor Laboratory (1993).  
Marshall, W.S., and Garthwaite, M.H., *Science* 259:1564  
Massing, J., *Methods in Enzymol.* 201:20 (1983).  
Michele, et al., *International Symposium on Viral*  
Miller, J. H., *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory (1972).  
10 Hepatitis.  
Mizller, J. H., *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory (1989).  
Moss, B., et al., *CURRENT Protocols in Molecular Biology* (1989).  
Mueller, K.B., U.S. Patent No. 4,683,202, issued 28 July 1987.  
Mueller, K.B., U.S. Patent No. 4,683,195, issued 28 July 1987.  
Osikowski, G., et al., *Clin. Chem.* 36:1586 (1990).  
Patterson, J.L., and Fernandes-Larsson, R., Rev.  
Infect. Dis. 12:113 (1990).  
Pearson, W.R., *Methods in Enzymology* 183:63-  
25 (1988).  
Perrath, J., *Protein Exp. and Purif.* 3:263 (1992).  
Pritchard, C.G., and Stefano, J.E., *Ann. Biol. Chem.* 98 (1990).  
Reichard, O., et al., *Lancet* 337:1058 (1991).  
Reiley, P.R., et al., *BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL* (1992).  
Reyes, G., et al., *Science*, 247:1335 (1990).  
Reyes, G., et al., *Molecular and Cellular Probes* 5:473-481 (1991).  
Roberts, N.A., et al., *Science* 248:358 (1990).

- Sanger, et al., *Proc. Natl. Acad. Sci.* 74:5463 (1977).
- 5 Sambrook, J., et al., In MOLECULAR CLONING: A LABORATORY MANUAL, Cold Spring Harbor Laboratory Press, Vol. 2 (1989).
- Saiki, R.K., et al., *Science* 239:487-491 (1988).
- Scharf, S. J., et al., *Science* 233:1076 (1986).
- Schuler, G.D., et al., *Proteins: Struc., Func. and Genet.* 9:180 (1989).
- Scott, J.K., and Smith, G.P., *Science* 249:386 (1990).
- 10 Scott, J.K., et al., *Proc. Natl. Acad. Sci. USA* 89:5398 (1992).
- Smith, D.B., et al., *Gene* 67:31 (1988).
- Smith, J.P., *Curr. Opin. Biotechnol.* 2:668 (1991).
- Sreenivasan, M.A., et al., *J. Gen. Virol.* 65:1005 (15) (1984).
- Tam, A., et al., *Virology* 185:120 (1991).
- Tam, J.P., *Proc. Natl. Acad. Sci. USA* 85:5409 (1988).
- Tonkinson, J.L., and Stein, C.A., *Antiviral Chem. and Chemother.* 4(4):193-200 (1993).
- 20 Ulmer, et al., *Science* 259:1745 (1993).
- Urdea, M., *Clin. Chem.* 39:725 (1993).
- Urdea, M., et al., *AIDS* 7:S11 (1993).
- Wages, J.M., et al., *Amplifications* 10:1-6 (1993).
- Walker, G.T., *PCR Methods Appl.* 3:1-6 (1993).
- 25 Wang, A.M., et al. in PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS (M.A. Innis, et al., eds.) Academic Press (1990).
- Wang, B., et al., *Proc. Natl. Acad. Sci. USA* 90:4156 (1993).
- 30 Whetsell, A.J., et al., *J. Clin. Micro.* 30:845 (1992).
- Wolf, J.A., et al., *Nature* 247:1465 (1990).
- Vacca, J.P., et al., *PNAS* 91:4096 (1994).
- VanGemen, B., et al., *J. Virol. Methods* 43:177 (35) (1993).
- Valenzuela, P., et al., *Nature* 298:344 (1982).

- Valenzuela, P., et al., in HEPATITIS B, eds. I.  
Millman, et al., Plenum Press, pages 225-236 (1984).  
Yarbrough, et al., J. Virol. 65:5790 (1991).  
Yoshio, T., et al., U.S. Patent No. 4,849,350, issued  
5 July 18, 1989.

#### BACKGROUND OF THE INVENTION

Viral hepatitis resulting from a virus other than hepatitis A virus (HAV) and hepatitis B virus (HBV) has 10 been referred to as non-A, non-B hepatitis (NANBH). NANBH can be further defined based on the mode of transmission of an individual type, for example, enteric versus parenteral.

One form of NANBH, known as enterically transmitted 15 NANBH or ET-NANBH, is contracted predominantly in poor-sanitation areas where food and drinking water have been contaminated by fecal matter. The molecular cloning of the causative agent, referred to as the hepatitis E virus (HEV), has recently been described (Reyes et al., 1990; 20 Tam et al.).

A second form of NANB, known as parenterally transmitted NANBH, or PT-NANBH, is transmitted by parenteral routes, typically by exposure to blood or blood products. The rate of this hepatitis varied by (i) locale, (ii) 25 whether ALT testing was done in blood banks, and (iii) elimination of high-risk patients for AIDS. Approximately 10% of transfusions caused PT-NANBH infection and about half of those went on to a chronic disease state (Dienstag). After implementation of anti-HCV testing, HCV 30 seroconversion per unit transfused was decreased to less than 1% among heart surgery patients (Alter).

Human plasma samples documented as having produced post-transfusion NANBH in human recipients have been used successfully to produce PT-NANBH infection in chimpanzees 35 (Bradley). RNA isolated from infected chimpanzee plasma has been used to construct cDNA libraries in an expression vector for immunoscreening with serum from human subjects

- with chronic PT-NANBH infection. This procedure identified a PT-NANBH specific cDNA clone and the viral sequence was then used as a probe to identify a set of overlapping fragments making up 7,300 contiguous basepairs of a PT-NANBH viral agent. The sequenced viral agent has been named the hepatitis C virus (HCV) (for example, the sequence of HCV is presented in EPO patent application 88310922.5, filed 11/18/88). The full-length sequence (~ 9,500 nt) of HCV is now available.
- 10 Primate transmission studies conducted at the Centers for Disease Control (CDC; Phoenix, AZ, 1973-1975; 1978-1983) originally provided substantial evidence for the existence of multiple agents of non-A, non-B hepatitis (NANBH): the primary agents associated with the majority
- 15 of cases of NANBH are now recognized to be HCV and HEV (see above), for PT-NANBH and ET-NANBH, respectively. Later epidemiologic studies conducted at the CDC (Atlanta, GA, 1989-present) using both research (prototype) and commercial tests for anti-HCV antibody showed that
- 20 approximately 20% of all community-acquired NANBH was also non-C. Further testing of these samples for the presence of HEV (co-owned, co-pending U.S. Application Serial No. 07/372,711, filed 28 June 1989, herein incorporated by reference) have indicated that these cases of community-
- 25 acquired non-A, non-B, non-C hepatitis were also non-E.
- Liver biopsy specimens, sera and plasma of Sentinel County patients (study of Drs. Miriam Alter and Kris Krawczynski) also showed that many *bona fide* cases of NANBH were also non-C hepatitis (serologically and by
- 30 Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR; Kawasaki, et al.; Wang, et al., 1990) negative for all markers of HCV infection) developed subsequently into chronic hepatitis with presentation of chronic persistent hepatitis (CPH) or chronic active hepatitis (CAH)
- 35 consistent with a viral infection.

**SUMMARY F THE INVENTION**

The present invention describes polypeptide antigens encoded by the reverse-frame of a selected virus having an RNA genome, where the polypeptide antigen is specifically immunoreactive with serum infected with the selected RNA virus. Reverse-frames are defined as open reading frames that are transcribed and translated in the opposite direction to the major known reading frames for the virus.

In one embodiment of the present invention the selected virus is a single, positive strand RNA virus. Exemplary viruses of this group are Hepatitis G Virus, also disclosed herein, and Hepatitis C Virus.

In another aspect, the present invention includes a method for detecting serum infected with a virus having an RNA genome. In this method, serum from a test subject is reacted with a reverse-frame polypeptide antigen. The polypeptide antigen is then examined for the presence of bound antibody. Alternatively, antibodies against the reverse-frame polypeptide antigen may be used to detect the presence of the reverse-frame polypeptide antigen in a sample.

In one embodiment of the detection method, a polypeptide antigen is attached to a solid support. The serum is then exposed to the polypeptide antigen/support followed by addition of a reporter-labelled anti-human antibody. The polypeptide antigen/support is then examined to detect the presence of reporter-labelled antibody bound to the polypeptide antigen/support.

The invention also includes antibodies directed against reverse-frame polypeptide antigens, including monoclonal antibodies and substantially isolated preparations of polyclonal antibodies.

Further, the invention includes diagnostic kits containing the above described reverse-frame polypeptide antigens and/or antibodies against these polypeptide antigens.

In another embodiment, the present invention includes a method of identifying a polypeptide antigen that is specifically immunoreactive with antibodies against a selected virus having an RNA genome. In the method,

5 polynucleotide sequences corresponding to the coding sequences for identifiable viral proteins are determined for the selected virus. A second polynucleotide sequence complementary to the first polynucleotide (encoding identifiable viral protein(s)) is examined for the

10 presence of an open reading frame (ORF). The immunological properties of the polypeptide encoded by the open reading frame are then examined to determine if the polypeptide is specifically immunoreactive with antibodies (e.g., infected serum) against the virus.

15 In one embodiment, the first polynucleotide is the genomic strand of a single, positive strand RNA virus (for example, HCV) that encodes a polyprotein.

Also, the following step can be included in the method of identifying a polypeptide antigen. Reverse-frames from a number of variants can be compared to determine the reverse-frame coding sequences that are conserved between variants. These conserved reverse-frame polypeptides are then evaluated for their antigenic properties.

25

These and other objects and features of the invention will be more fully appreciated when the following detailed description of the invention is read in conjunction with the accompanying drawings.

30

#### BRIEF DESCRIPTION OF THE FIGURES

Figure 1: the relationship of the SEQ ID NO:14 open reading frame to the 470-20-1 clone.

35 Figure 2: shows an exemplary protein profile from gradient fractions eluted from a glutathione affinity column.

Figure 3: shows an exemplary Sodium dodecyl sulfate polyacrylamide gel electrophoresis analysis of fraction samples from Figure 2.

5 Figure 4A: shows an exemplary protein profile from gradient fractions eluted from an anion exchange column.

Figures 4B and 4C: show exemplary Sodium dodecyl sulfate polyacrylamide gel electrophoresis analysis of fraction samples from Figure 4A.

10 Figures 5A and 5B: amino acid alignments of HGV with two other members of Flaviviridae family -- Hog Cholera Virus and Hepatitis C Virus.

Figure 6 shows a map of a portion of the vector pGEX-Hisb-GE3-2, a bacterial expression plasmid carrying an HGV epitope.

15 Figures 7A to 7D show the results of Western blot analysis of the purified HGV GE3-2 protein.

Figures 8A to 8D show the results of Western blot analysis of the purified HGV Y5-10 antigen.

20 Figures 9A to 9D show the results of Western blot analysis of the following antigens: Y5-5, GE3-2 and Y5-10.

Figure 10: shows the relative positions of two exemplary reverse open reading frame antigens.

25 Figures 11A, 11B and 11C show a multiple sequence alignment for the K3 clones.

#### DETAILED DESCRIPTION OF THE INVENTION

##### I. DEFINITIONS

The terms defined below have the following meaning  
30 herein:

1. "nonA/nonB/nonC/nonD/nonE hepatitis viral agent {N-(ABCDE)}," herein provisionally designated HGV, means a virus, virus type, or virus class which (i) is transmissible in some primates, including, mystax, chimpanzees or humans, (ii) is serologically distinct from hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), hepatitis D virus, and hepatitis

E (HEV) (although HGV may co-infect a subject with these viruses), and (iii) is a member of the virus family Flaviviridae.

2. "HGV variants" are defined as viral isolates that  
5 have at least about 40%, preferably 55%, more preferably  
70%, or most preferably 80% global sequence homology, that  
is, sequence identity over a length (comparable to SEQ ID  
NO:14) of the viral genome polynucleotide sequence, to the  
HGV polynucleotide sequences disclosed herein.

10 "Sequence homology" is determined essentially as follows. Two polynucleotide sequences of the same length (preferably, the entire viral genome) are considered to be homologous to one another, if, when they are aligned using the ALIGN program, over 40%, or preferably 55%, more  
15 preferably 70%, or most preferably 80% of the nucleic acids in the highest scoring alignment are identically aligned using a ktup of 1, the default parameters and the default PAM matrix.

The ALIGN program is found in the FASTA version 1.7  
20 suite of sequence comparison programs (Pearson, et al.,  
1988; Pearson, 1990; program available from William R.  
Pearson, Department of Biological Chemistry, Box 440,  
Jordan Hall, Charlottesville, VA).

In determining whether two viruses are "highly  
25 homologous" to each other, the complete sequence of all the viral proteins (or the polyprotein) for one virus are optimally, globally aligned with the viral proteins or polyprotein of the other virus using the ALIGN program of the above suite using a ktup of 1, the default parameters  
30 and the default PAM matrix. Regions of dissimilarity or similarity are not excluded from the analysis.

Differences in lengths between the two sequences are considered as mismatches. Alternatively, viral structural protein regions are typically used to determine  
35 relatedness between viral isolates. Highly homologous viruses have over 40%, or preferably 55%, more preferably

70%, or most preferably 80% global polypeptide sequence identity.

3. Two nucleic acid fragments are considered to be "selectively hybridizable" to an HGV polynucleotide, if
- 5 they are capable of specifically hybridizing to HGV or a variant thereof (e.g., a probe that hybridizes to HGV nucleic acid but not to polynucleotides from other members of the virus family Flaviviridae) or specifically priming a polymerase chain reaction: (i) under typical
- 10 hybridization and wash conditions, as described, for example, in Maniatis, et al., pages 320-328, and 382-389, or (ii) using reduced stringency wash conditions that allow at most about 25-30% basepair mismatches, for example: 2 × SSC, 0.1% SDS, room temperature twice, 30
- 15 minutes each; then 2 × SSC, 0.1% SDS, 37°C. once, 30 minutes; then 2 × SSC room temperature twice, 10 minutes each, or (iii) selecting primers for use in typical polymerase chain reactions (PCR) under standard conditions (for example, in Saiki, R.K., et al.), which result in
- 20 specific amplification of sequences of HGV or its variants.

Preferably, highly homologous nucleic acid strands contain less than 20-30% basepair mismatches, even more preferably less than 5-20% basepair mismatches. These

25 degrees of homology can be selected by using wash conditions of appropriate stringency for identification of clones from gene libraries (or other sources of genetic material), as is well known in the art.

4. An "HGV polynucleotide," as used herein, is
- 30 defined as follows. For polynucleotides greater than about 100 nucleotides, HGV polynucleotides encompass polynucleotide sequences encoded by HGV variants and homologous sequences as defined in "2" above. For polynucleotides less than about 100 nucleotides in length,
- 35 HGV polynucleotide encompasses sequences that selectively hybridizes to sequences of HGV or its variants. Further,

HGV polynucleotides include polynucleotides encoding HGV polypeptides (see below).

- The term "polynucleotide" as used herein refers to a polymeric molecule having a backbone that supports bases capable of hydrogen bonding to typical nucleic acids, where the polymer backbone presents the bases in a manner to permit such hydrogen bonding in a sequence specific fashion between the polymeric molecule and a typically nucleic acid (e.g., single-stranded DNA). Such bases are typically inosine, adenosine, guanosine, cytosine, uracil and thymidine. Numerous polynucleotide modifications are known in the art, for example, labels, methylation, and substitution of one or more of the naturally occurring nucleotides with an analog.
- 15        Polymeric molecules include double and single stranded RNA and DNA, and backbone modifications thereof, for example, methylphosphonate linkages. Further, such polymeric molecules include alternative polymer backbone structures such as, but not limited to, polyvinyl backbones (Pitha, 1970a/b), morpholino backbones (Summerton, et al., 1992, 1993). A variety of other charged and uncharged polynucleotide analogs have been reported. Numerous backbone modifications are known in the art, including, but not limited to, uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, and carbamates), charged linkages (e.g., phosphorothioates and phosphorodithioates). In addition linkages may contain the following exemplary modifications: pendant moieties, such as, proteins (including, for example, nucleases, toxins, antibodies, signal peptides and poly-L-lysine); intercalators (e.g., acridine and psoralen), chelators (e.g., metals, radioactive metals, boron and oxidative metals), alkylators, and other modified linkages (e.g., alpha anomeric nucleic acids).

5. An "HGV polypeptide" is defined herein as any peptide homologous to an HGV polypeptide. "Homology,"

as used herein, is defined as follows. In one embodiment, a polypeptide is homologous to an HGV polypeptide if it is encoded by nucleic acid that selectively hybridizes to sequences of HGV or its variants.

- 5        In another embodiment, a polypeptide is homologous to an HGV polypeptide if it is encoded by HGV or its variants, as defined above, polypeptides of this group are typically larger than 15, preferable 25, or more preferable 35, contiguous amino acids. Further, for
- 10      polypeptides longer than about 60 amino acids, sequence comparisons for the purpose of determining "polypeptide homology" are performed using the local alignment program LALIGN. The polypeptide sequence, is compared against the HGV amino acid sequence or any of its variants, as defined
- 15      above, using the LALIGN program with a ktup of 1, default parameters and the default PAM.

Any polypeptide with an optimal alignment longer than 60 amino acids and greater than 65%, preferably 70%, or more preferably 80% of identically aligned amino acids is

- 20      considered to be a "homologous polypeptide." The LALIGN program is found in the FASTA version 1.7 suite of sequence comparison programs (Pearson, et al., 1988; Pearson, 1990; program available from William R. Pearson, Department of Biological Chemistry, Box 440, Jordan Hall, 25 Charlottesville, VA).

6. A polynucleotide is "derived from" HGV if it has the same or substantially the same basepair sequence as a region of an HGV genome, cDNA of HGV or complements thereof, or if it displays homology as noted under "2",
- 30      "3" or "4" above.

A polypeptide is "derived from" HGV if it is (i) encoded by an open reading frame of an HGV polynucleotide, or (ii) displays homology to HGV polypeptides as noted under "2" and "5" above, or (iii) is specifically

- 35      immunoreactive with HGV positive sera.

7. "Substantially isolated" and "purified" are used in several contexts and typically refer to at least

partial purification of an HGV virus particle, component (e.g., polynucleotide or polypeptide), or related compound (e.g., anti-HGV antibodies) away from unrelated or contaminating components (e.g., serum cells, proteins,

5 non-HGV polynucleotides and non-anti-HGV antibodies).

Methods and procedures for the isolation or purification of compounds or components of interest are described below (e.g., affinity purification of fusion proteins and recombinant production of HGV polypeptides).

10 8. In the context of the present invention, the phrase "nucleic acid sequences," when referring to sequences which encode a protein, polypeptide, or peptide, is meant to include degenerative nucleic acid sequences which encode homologous protein, polypeptide or peptide sequences as well as the disclosed sequence.

15 9. An "epitope" is the antigenic determinant defined as the specific portion of an antigen with which the antigen binding portion of a specific antibody interacts.

10. An antigen or epitope is "specifically immunoreactive" with HGV positive sera when the epitope/antigen binds to antibodies present in the HGV infected sera but does not bind to antibodies present in the majority (greater than about 90%, preferably greater than 95%) of sera from individuals who are not or have not been infected with HGV. "Specifically immunoreactive" antigens or epitopes may also be immunoreactive with monoclonal or polyclonal antibodies generated against specific HGV epitopes or antigens.

An antibody or antibody composition (e.g., polyclonal antibodies) is "specifically immunoreactive" with HGV when the antibody or antibody composition is immunoreactive with an HGV antigen but not with HAV, HBV, HCV, HDV or HEV antigens. Further, "specifically immunoreactive antibodies" are not immunoreactive with antigens typically present in normal sera obtained from subjects not infected with or exposed to HGV, HAV, HBV, HCV, HDV or HEV.

## II. ISOLATION OF HGV ASSOCIATED SEQUENCES.

As one approach toward identifying clones containing HGV sequences, a cDNA library was prepared from infected-HGV sera in the expression vector lambda gt11 (Example 1).  
5 Polynucleotide sequences were then selected for the expression of peptides which are immunoreactive with serum PNF 2161. PNF 2161 was believed to contain an etiologic agent of NANBH other than HCV. First round screening was typically performed using the PNF 2161 serum (used to  
10 generate the phage library). It is also possible to screen with other suspected N-(ABCDE) sera.

Recombinant proteins identified by this approach provide candidates for peptides which can serve as substrates in diagnostic tests. Further, the nucleic acid  
15 coding sequences identified by this approach serve as useful hybridization probes for the identification of additional HGV coding sequences.

The sera described above were used to generate cDNA libraries in lambda gt11 (Example 1). In the method  
20 illustrated in Example 1, infected serum was precipitated in 8% PEG without dilution, and the libraries were generated from the resulting pelleted virus. Sera from infected human sources were treated in the same fashion.

As an advantageous alternative to PEG precipitation,  
25 ultracentrifugation can be used to pellet particulate agents from infected sera or other biological specimens. To isolate viral particles from which nucleic acids could be extracted, serum, ranging up to 2 ml, is diluted to approximately 10 ml with PBS, spun at 3K for 10 minutes,  
30 and the supernatant is centrifuged for a minimum of 2 hours at 40,000 rpm (approximately 110,000 x g) in a Ti70.1 rotor (Beckman Instruments, Fullerton, CA) at 4°C. The supernatant is then aspirated and the pellet extracted by standard nucleic acid extraction techniques.

35 cDNA libraries were generated using random primers in reverse transcription reactions with RNA extracted from pelleted sera as starting material. The resulting

molecules were ligated to Sequence Independent Single Primer Amplification (SISPA; Reyes, et al., 1991) linker primers and expanded in a non-selective manner, and then cloned into a suitable vector, for example, lambda gt11,  
5 for expression and screening of peptide antigens.

Alternatively, the lambda gt10 vector may also be used.

Lambda gt11 is a particularly useful expression vector which contains a unique EcoRI insertion site 53 base pairs upstream of the translation termination codon  
10 of the  $\beta$ -galactosidase gene. Thus, an inserted sequence is expressed as a  $\beta$ -galactosidase fusion protein which contains the N-terminal portion of the  $\beta$ -galactosidase gene product, the heterologous peptide, and optionally the C-terminal region of the  $\beta$ -galactosidase peptide (the C--  
15 terminal portion being expressed when the heterologous peptide coding sequence does not contain a translation termination codon).

This vector also produces a temperature-sensitive repressor (cI857) which causes viral lysogeny at permissive temperatures, e.g., 32°C, and leads to viral lysis at elevated temperatures, e.g., 42°C. Advantages of this vector include: (1) highly efficient recombinant clone generation, (2) ability to select lysogenized host cells on the basis of host-cell growth at permissive, but not  
25 non-permissive, temperatures, and (3) production of recombinant fusion protein. Further, since phage containing a heterologous insert produces an inactive  $\beta$ -galactosidase enzyme, phage with inserts are typically identified using a colorimetric substrate conversion reaction employing  $\beta$ -galactosidase.  
30

Example 1 describes the preparation of a cDNA library for the N-(ABCDE) hepatitis sera PNF 2161. The library was immunoscreened using PNF 2161 (Example 3). A number of lambda gt11 clones were identified which were  
35 immunoreactive. Immunopositive clones were plaque-purified and their immunoreactivity retested. Also, the

immunoreactivity of the clones with normal human sera was also tested.

These clones were also examined for the "exogenous" nature of the cloned insert sequence. This basic test 5 establishes that the cloned fragment does not represent a portion of human or other potentially contaminating nucleic acids (e.g., *E. coli*, *S. cerevisiae* and mitochondrial). The clone inserts were isolated by EcoRI digestion following polymerase chain reaction 10 amplification. The inserts were purified then radiolabelled and used as hybridization probes against membrane bound normal human DNA, normal mystax DNA and bacterial DNA (control DNAs) (Example 4A).

Clone 470-20-1 (PNF2161 cDNA source) was one of the 15 clones isolated by immunoscreening with the PNF 2161 serum. The clone was not reactive with normal human sera. The clone has a large open reading frame (203 base pairs; SEQ ID NO:3), in-frame with the  $\beta$ -galactosidase gene of the lambda-gt11 vector. The clone is exogenous by genomic 20 DNA hybridization analysis and genomic PCR analysis, using human, yeast and *E. coli* genomic DNAs (Example 4B).

The sequence was present in PNF2161 serum as determined by RT-PCR (Example 4C). RT-PCR of serially diluted PNF 2161 RNA suggested at least about  $10^5$  copies of 470-20-25 1 specific sequence per ml. The sequence was also detected in sucrose density gradient fractions at densities consistent with the sequence banding in association with a virus-like particle (Example 5).

Bacterial lysates of *E. coli* expressing a second 30 clone, clone 470-exp1, (SEQ ID NO:28) were also shown to be specifically immunoreactive with PNF 2161 serum at comparable levels to clone 470-20-1. The coding sequence of 470-exp1 was flanked by termination codons (based on sequence comparisons to SEQ ID NO:14, also see Figure 1) 35 and had an internal methionine.

Further sequences (SEQ ID NO:14) adjacent to clone 470-20-1 were obtained by anchor polymerase chain reaction

(Anchor PCR) using primers from clone 470-20-1 (Example 6). In this case a PNF 2161 2-cDNA source library was used as template, where the cDNA/complement double-stranded DNA products were ligated to lambda arms, but the 5 mixture was not packaged.

470-20-1 specific primers were used in amplification reactions with SISPA-amplified PNF 2161 cDNA as a template (Example 4). The identity of the amplified DNA fragments were confirmed by (i) size and (ii) hybridization with a 10 470-20-1 specific oligonucleotide probe (SEQ ID NO:16). The 470-20-1 specific signal was detected in cDNA amplified by PCR from SISPA-amplified PNF 2161, demonstrating the presence of the 470-20-1 sequences in the source material.

15 The 470-20-1 specific primers were also used in amplification reactions with the following RNA sources as substrate: normal mystax liver RNA, normal tamarin (*Sanguins laboriatis*) liver RNA, and MY131 liver RNA (Example 4). The results from these experiments demonstrate the 470-20-1 sequences are present in the parent 20 serum sample (PNF 2161) and in an RNA liver sample from an animal challenged with the PNF 2161 sample (MY131). Both normal control RNAs were negative for the presence of 470-20-1 sequences.

25 Further, PNF 2161 serum and other cloning source or related source materials were directly tested by PCR using primers from selected cloned sequences. Specific amplification products were detected by hybridization to a specific oligonucleotide probe 470-20-1-152F (SEQ ID NO:16). A specific signal was reproducibly detected in 30 multiple extracts of PNF 2161, with the 470-20-1 specific primers.

The disease association between HGV and liver disease is further supported by the data presented in Example 4F. 35 Sera from hepatitis patients and from blood donors with abnormal liver function were assessed for the presence of HGV by RT-PCR screening, using HGV specific primers. HGV

specific sequence were detected in 6/152 of these sera samples. No HGV positives were detected among the control samples (n = 11).

The results presented above indicate the isolation of  
5 a viral agent associated with N-(ABCDE) viral infection of liver (i.e., hepatitis) and/or infection, and resulting disease, of other tissue and cell types. Cloning of further HGV isolates (JC, BG34, T55806 and EB20) is described in Example 15.

10

### III. FURTHER CHARACTERIZATION OF HGV RECOMBINANT ANTIGENS.

#### A. SCREENING RECOMBINANT LIBRARIES.

Further candidate HGV antigens can be obtained from the libraries of the present invention using the screening  
15 methods described above. The cDNA library described above has been deposited with the American Type Culture Collection, 12301 Parklawn Dr., Rockville, MD, 20852, and has been assigned the following designation: PNF 2161 cDNA source, ATCC 75268.

20 A second PNF 2161 cDNA library has been generated essentially as described for the first PNF 2161 cDNA library, except that second PNF 2161 cDNA source library was ligated to lambda gt11 arms but was not packaged. This non-packaged library was used to obtain the extension  
25 clones described below. A packaged version of this second library (PNF 2161 2-cDNA source library) has been deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD, 20852, and has been assigned the following designation: PNF 2161 2-cDNA  
30 source, ATCC 75837.

In addition to the recombinant libraries generated above, other recombinant libraries from N-(ABCDE) hepatitis sera can likewise be generated and screened as described herein.

35

B. EPITOPE MAPPING, CROSS HYBRIDIZATION AND ISOLATION OF GENOMIC SEQUENCES.

Antigen encoding DNA fragments can be identified by (i) immunoscreening, as described above, or (ii) computer analysis of coding sequences (e.g., SEQ ID NO:14) using an algorithm (such as, "ANTIGEN," Intelligenetics, Mountain View, CA) to identify potential antigenic regions. An antigen-encoding DNA fragment can be subcloned. The subcloned insert can then be fragmented by partial DNase I digestion to generate random fragments or by specific restriction endonuclease digestion to produce specific subfragments. The resulting DNA fragments can be inserted into the lambda gt11 vector and subjected to immuno-screening in order to provide an epitope map of the cloned insert.

In addition, the DNA fragments can be employed as probes in hybridization experiments to identify overlapping HGV sequences, and these in turn can be further used as probes to identify a set of contiguous clones. The generation of sets of contiguous clones allows the elucidation of the sequence of the HGV's genome.

Any of the above-described clone sequences (e.g., derived from SEQ ID NO:14 or clone 470-20-1) can be used to probe the cDNA and DNA libraries, generated in a vector such as lambda gt10 or "LAMBDA ZAP II" (Stratagene, San Diego, CA). Specific subfragments of known sequence may be isolated by polymerase chain reaction or after restriction endonuclease cleavage of vectors carrying such sequences. The resulting DNA fragments can be used as radiolabelled probes against any selected library. In particular, the 5' and 3' terminal sequences of the clone inserts are useful as probes to identify additional clones.

Further, the sequences provided by the 5' end of cloned inserts are useful as sequence specific primers in first-strand cDNA or DNA synthesis reactions (Maniatis et al.; Scharf et al.). For example, specifically primed PNF

2161 cDNA and DNA libraries can be prepared by using specific primers derived from SEQ ID NO:14 on PNF 2161 nucleic acids as a template. The second-strand of the new cDNA is synthesized using RNase H and DNA polymerase I.

- 5 The above procedures identify or produce DNA/cDNA molecules corresponding to nucleic acid regions that are 5' adjacent to the known clone insert sequences. These newly isolated sequences can in turn be used to identify further flanking sequences, and so on, to identify the
- 10 sequences composing the entire genome for HGV. As described above, after new HGV sequences are isolated, the polynucleotides can be cloned and immunoscreened to identify specific sequences encoding HGV antigens.

Extension clone sequences (SEQ ID NO:14), containing further sequences of interest, were obtained for clone PNF 470-20-1 (SEQ ID NO:3) using the "Anchor PCR" method described in Example 6. Briefly, the strategy consists of ligating PNF 2161 SISPA cDNA to lambda gt11 arms and amplifying the ligation reaction with a gt11-specific primer and one of two 470-20-1 specific primers.

- 20
  - 25
- The amplification products are electrophoretically separated, transferred to filters and the DNA bound to the filters is probed with a 470-20-1 specific probe. Bands corresponding to hybridization positive band signals were gel purified, cloned and sequenced.

#### C. PREPARATION OF ANTIGENIC POLYPEPTIDES AND ANTIBODIES.

The recombinant peptides of the present invention can be purified by standard protein purification procedures which may include differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel electrophoresis and affinity chromatography.

In one embodiment of the present invention, the polynucleotide sequences of the antigens of the present invention have been cloned in the plasmid p-GEX (Example 7A) or various derivatives thereof (pGEX-GLI). The plas-

mid pGEX (Smith, et al., 1988) and its derivatives express the polypeptide sequences of a cloned insert fused in-frame to the protein glutathione-S-transferase (sj26). In one vector construction, plasmid pGEX-hisB, an amino acid sequence of 6 histidines is introduced at the carboxy terminus of the fusion protein.

The various recombinant pGEX plasmids can be transformed into appropriate strains of *E. coli* and fusion protein production can be induced by the addition of IPTG (isopropyl-thio galactopyranoside) as described in Example 7A. Solubilized recombinant fusion protein can then be purified from cell lysates of the induced cultures using glutathione agarose affinity chromatography (Example 7A).

Insoluble fusion protein expressed by the plasmid pGEX-hisB can be purified by means of immobilized metal ion affinity chromatography (Porath) in buffers containing 6M Urea or 6 M guanidinium isothiocyanate, both of which are useful for the solubilization of proteins. Alternatively insoluble proteins expressed in pGEX-GLI or derivatives thereof can be purified using combinations of centrifugation to remove soluble proteins followed by solubilization of insoluble proteins and standard chromatographic methodologies, such as ion exchange or size exclusion chromatography, and other such methods are known in the art.

In the case of  $\beta$ -galactosidase fusion proteins (such as those produced by lambda gt11 clones) the fused protein can be isolated readily by affinity chromatography, by passing cell lysis material over a solid support having surface-bound anti- $\beta$ -galactosidase antibody. For example, purification of a  $\beta$ -galactosidase/fusion protein, derived from 470-20-1 coding sequences, by affinity chromatography is described in Example 7B.

Also included in the invention is an expression vector, such as the lambda gt11 or pGEX vectors described above, containing HGV coding sequences and expression control elements which allow expression of the coding

regions in a suitable host. The control elements generally include a promoter, translation initiation codon, and translation and transcription termination sequences, and an insertion site for introducing the insert into the  
5 vector.

The DNA encoding the desired antigenic polypeptide can be cloned into any number of commercially available vectors to generate expression of the polypeptide in the appropriate host system. These systems include, but are  
10 not limited to, the following: baculovirus expression (Reilly, et al.; Beames, et al.; Pharmingen; Clontech, Palo Alto, CA), vaccinia expression (Moss, et al.), expression in bacteria (Ausubel, et al.; Clontech), expression in yeast (Goeddel; Guthrie and Fink), expression in  
15 mammalian cells (Clontech; Gibco-BRL, Ground Island, NY). These recombinant polypeptide antigens can be expressed directly or as fusion proteins. A number of features can be engineered into the expression vectors, such as leader sequences which promote the secretion of the expressed  
20 sequences into culture medium. The recombinantly produced HGV polypeptide antigens are typically isolated from lysed cells or culture media. Purification can be carried out by methods known in the art including salt fractionation, ion exchange chromatography, and affinity chromatography.  
25 Immunoaffinity chromatography can be employed using antibodies generated based on the HGV antigens identified by the methods of the present invention.

HGV polypeptide antigens may also be isolated from HGV particles (see below).

30 Continuous antigenic determinants of polypeptides are generally relatively small, typically 6 to 10 amino acids in length. Smaller fragments have been identified as antigenic regions, for example, in conformational epitopes. HGV polypeptide antigens are identified as  
35 described above. The resulting DNA coding regions of either strand can be expressed recombinantly either as fusion proteins or isolated polypeptides. In addition,

amino acid sequences can be conveniently chemically synthesized using commercially available synthesizer (Applied Biosystems, Foster City, CA) or "PIN" technology (Applied Biosystems).

- 5        In another embodiment, the present invention includes mosaic proteins that are composed of multiple epitopes. An HGV mosaic polypeptide typically contains at least two epitopes of HGV, where the polypeptide substantially lacks amino acids normally intervening between the epitopes in
- 10      the native HGV coding sequence. Synthetic genes (Crea; Yoshio et al.; Eaton et al.) encoding multiple, tandem epitopes can be constructed that will produce mosaic proteins using standard recombinant DNA technology using polypeptide expression vector/host system described above.
- 15      Further, multiple antigen peptides can be synthesized chemically by methods described previously (Tam, J.P., 1988; Briand et al.). For example, a small immuno-
- 20      logically inert core matrix of lysine residues with  $\alpha$ - and  $\epsilon$ - amino groups can be used to anchor multiple copies of the same or different synthetic peptides (typically 6-15 residues long) representing epitopes of interest. Mosaic proteins or multiple antigen peptide antigens give higher sensitivity and specificity in immunoassays due to the signal amplification resulting from distribution of
- 25      multiple epitopes.

Antigens obtained by any of these methods can be used for antibody generation, diagnostic tests and vaccine development.

- In another aspect, the invention includes specific
- 30      antibodies directed against the polypeptide antigens of the present invention. Antigens obtained by any of these methods may be directly used for the generation of antibodies or they may be coupled to appropriate carrier molecules. Many such carriers are known in the art and
- 35      are commercially available (e.g., Pierce, Rockford IL). Typically, to prepare antibodies, a host animal, such as a rabbit, is immunized with the purified antigen or fused

protein antigen. Hybrid, or fused, proteins may be generated using a variety of coding sequence derived from other proteins, such as glutathione-S-transferase or  $\beta$ -galactosidase. The host serum or plasma is collected 5 following an appropriate time interval, and this serum is tested for antibodies specific against the antigen. Example 8 describes the production of rabbit serum antibodies which are specific against the 470-20-1 antigen in the Sj26/470-20-1 hybrid protein. These techniques are 10 equally applicable to all immunogenic sequences derived from HGV, including, but not limited to, those derived from the coding sequence presented as SEQ ID NO:14.

The gamma globulin fraction or the IgG antibodies of immunized animals can be obtained, for example, by use of 15 saturated ammonium sulfate precipitation or DEAE Sephadex chromatography, affinity chromatography, or other techniques known to those skilled in the art for producing polyclonal antibodies.

Alternatively, purified antigen or fused antigen protein 20 may be used for producing monoclonal antibodies. Here the spleen or lymphocytes from an immunized animal are removed and immortalized or used to prepare hybridomas by methods known to those skilled in the art. To produce a human-derived hybridoma, a human lymphocyte donor is 25 selected. A donor known to be infected with a HGV may serve as a suitable lymphocyte donor. Lymphocytes can be isolated from a peripheral blood sample. Epstein-Barr virus (EBV) can be used to immortalize human lymphocytes or a suitable fusion partner can be used to produce human- 30 derived hybridomas. Primary *in vitro* sensitization with viral specific polypeptides can also be used in the generation of human monoclonal antibodies.

Antibodies secreted by the immortalized cells are screened to determine the clones that secrete antibodies 35 of the desired specificity, for example, by using the ELISA or Western blot method (Example 9; Ausubel et al.).

Using the antibodies of the present invention other antigenic peptides and epitopes can be isolated.

D. ELISA AND PROTEIN BLOT SCREENING.

When HGV antigens are identified, typically through plaque immunoscreening as described above, the antigens can be expressed and purified. The antigens can then be screened rapidly against a large number of suspected HGV hepatitis sera using alternative immunoassays, such as, 5 ELISAs or Protein Blot Assays (Western blots) employing the isolated antigen peptide. The antigen polypeptides fusion can be isolated as described above, usually by affinity chromatography to the fusion partner such as  $\beta$ -galactosidase or glutathione-S-transferase. Alternatively, 10 the antigen itself can be purified using antibodies generated against it (see below). 15

A general ELISA assay format is presented in Example 9. Harlow, et al., describe a number of useful techniques for immunoassays and antibody/antigen screening.

The purified antigen polypeptide or fusion polypeptide containing the antigen of interest, is attached to a solid support, for example, a multiwell polystyrene plate. Sera to be tested are diluted and added to the wells. After a period of time sufficient for the binding of 20 antibodies to the bound antigens, the sera are washed out of the wells. A labelled reporter antibody is added to each well along with an appropriate substrate: wells containing antibodies bound to the purified antigen polypeptide or fusion polypeptide containing the antigen are 25 detected by a positive signal. 30

A typical format for protein blot analysis using the polypeptide antigens of the present invention is presented in Example 9. General protein blotting methods are described by Ausubel, et al. In Example 9, the 470-20-35 1/sj26 fusion protein was used to screen a number of sera samples. The results presented in Example 9 demonstrate

that several different source N-(ABCDE) hepatitis sera are immunoreactive with the polypeptide antigen.

The results presented above demonstrate that the polypeptide antigens of the present invention can, by 5 these methods, be rapidly screened against panels of suspected HGV infected serum samples for the detection of HGV.

10 E. CELL CULTURE SYSTEMS, ANIMAL MODELS AND ISOLATION OF HGV.

HGV may be propagated in the animal model systems. Infectivity studies have been carried out in chimpanzees, cynomolgus monkey and four mystax subjects (Example 4G). These studies have yielded further information about HGV 15 infectivity in these animal models. The HGV described in the present specification have the advantage of being capable of infecting tamarins, cynomologous monkeys and chimpanzees.

Alternatively, primary hepatocytes obtained from 20 infected animals (chimpanzees, baboons, monkeys, or humans) can be cultured *in vitro*. A serum-free medium, supplemented with growth factors and hormones, has been described which permits the long-term maintenance of differentiated primate hepatocytes (Lanford, et al.; 25 Jacob, et al., 1989, 1990, 1991). In addition to primary hepatocyte cultures, immortalized cultures of infected cells may also be generated. For example, primary liver cultures may be fused to a variety of cells (like HepG2) to provide stable immortalized cell lines. Primary hepatocyte cell cultures may also be immortalized by introduction of oncogenes or genes causing a transformed phenotype. Such oncogenes or genes can be derived from a 30 number of sources known in the art including SV40, human cellular oncogenes and Epstein Barr Virus.

35 Further, the un-infected hepatocytes (e.g., primary or continuous hepatoma cell lines) may be infected by exposing the cells in culture to the HGV either as

- partially purified particle preparations (prepared, for example, from infected sera by differential centrifugation and/or molecular sieving) or in infectious sera. These infected cells can then be propagated and the virus
- 5   passaged by methods known in the art. Further, other cell types, such as lymphoid cell lines, may be useful for the propagation of HGV.

Protein similarity studies of HGV have detected amino acid regions similar to other viruses in the family

10   Flaviviridae. It is known that members of this family of viruses can be propagated in a variety of tissue culture systems (ATCC-Viruses catalogue, 1990). By analogy it is likely that HGV can be propagated in one or more of the following tissue culture systems: Hela cells, primary

15   hamster kidney cells, monkey kidney cells, vero cells, LLC-MK2 (rhesus monkey kidney cells), KB cells(human oral epidermoid carcinoma cells), duck embryo cells, primary sheep leptomeningeal cells, primary sheep choroid plexus cells, pig kidney cells, bovine embryonic kidney cells,

20   bovine turbinate cells, chick embryo cells, primary rabbit kidney cells, BHD-21 cells, or PK-13 cells.

In addition to expression of HGV, regions of HGV polynucleotide sequences, cDNA or in vitro transcribed RNA can be introduced by recombinant means into tissue culture

25   cells. Such recombinant manipulations allow the individual expression of individual components of the HGV.

RNA samples can be prepared from infected tissue or, in particular, from infected cell cultures. The RNA samples can be fractionated on gels and transferred to

30   membranes for hybridization analysis using probes derived from the cloned HGV sequences.

HGV particles may be isolated from infected sera, infected tissue, the above-described cell culture media, or the cultured infected cells by methods known in the

35   art. Such methods include techniques based on size fractionation (i.e., ultrafiltration, precipitation, sedimentation), using anionic and/or cationic exchange materials,

separation on the basis of density, hydrophilic properties, and affinity chromatography. During the isolation procedure the HGV can be identified (i) using the anti-HGV hepatitis associated agent antibodies of the 5 present invention, (ii) by using hybridization probes based on identified HGV nucleic acid sequences (e.g., Example 5) or (iii) by RT-PCR.

Antibodies directed against HGV can be used in purification of HGV particles through immunoaffinity chromatography (Harlow, et al.; Pierce). Antibodies directed 10 against HGV polypeptides or fusion polypeptides (such as 470-20-1) are fixed to solid supports in such a manner that the antibodies maintain their immunoselectivity. To accomplish such attachment of antibodies to solid support 15 bifunctional coupling agents (Pierce; Pharmacia, Piscataway, NJ) containing spacer groups are frequently used to retain accessibility of the antigen binding site of the antibody.

HGV particles can be further characterized by standard 20 procedures including, but not limited to, immunofluorescence microscopy, electron microscopy, Western blot analysis of proteins composing the particles, infection studies in animal and/or cell systems utilizing the partially purified particles, and sedimentation characteristics. The results presented in Example 5 suggest that the 25 viral particle of the present invention is more similar to an enveloped viral particle than to a non-enveloped viral particle.

HGV particles can be disrupted to obtain HGV genomes. 30 Disruption of the particles can be achieved by, for example, treatment with detergents in the presence of chelating agents. The genomic nucleic acid can then be further characterized. Characterization may include analysis of DNase and RNase sensitivity. The strandedness 35 (Example 4F) and conformation (e.g., circular) of the genome can be determined by techniques known in the art,

including visualization by electron microscopy and sedimentation characteristics.

The isolated genomes also make it possible to sequence the entire genome whether it is segmented or not, 5 and whether it is an RNA or DNA genome (using, for example RT-PCR, chromosome walking techniques, or PCR which utilizes primers from adjacent cloned sequences). Determination of the entire sequence of HGV allows genomic organization studies and the comparison of the HGV sequences to the coding and regulatory sequences of known 10 viral agents.

#### F. SCREENING FOR AGENTS HAVING ANTI-HGV HEPATITIS ACTIVITY.

The use of cell culture and animal model systems for 15 propagation of HGV provides the ability to screen for anti-hepatitis agents which inhibit the production of infectious HGV: in particular, drugs that inhibit the replication of HGV. Cell culture and animal models allow the evaluation of the effect of such anti-hepatitis drugs 20 on normal cellular functions and viability. Potential anti-viral agents (including, for example, small molecules, complex mixtures such as fungal extracts, and anti-sense oligonucleotides) are typically screened for anti-viral activity over a range of concentrations. The effect 25 on HGV replication and/or antigen production is then evaluated, typically by monitoring viral macromolecular synthesis or accumulation of macromolecules (e.g., DNA, RNA or protein). This evaluation is often made relative 30 to the effect of the anti-viral agent on normal cellular function (DNA replication, RNA transcription, general protein translation, etc.).

The detection of the HGV can be accomplished by many methods including those described in the present specification. For example, antibodies can be generated against 35 the antigens of the present invention and these antibodies used in antibody-based assays (Harlow, et al.) to identify and quantitate HGV antigens in cell culture. HGV antigens

can be quantitated in culture using competition assays: polypeptides encoded by the cloned HGV sequences can be used in such assays. Typically, a recombinantly produced HGV antigenic polypeptide is produced and used to generate 5 a monoclonal or polyclonal antibody. The recombinant HGV polypeptide is labelled using a reporter molecule. The inhibition of binding of this labelled polypeptide to its cognate antibody is then evaluated in the presence of samples (e.g., cell culture media or sera) that contain 10 HGV antigens. The level of HGV antigens in the sample is determined by comparison of levels of inhibition to a standard curve generated using unlabelled recombinant proteins at known concentrations.

The HGV sequences of the present invention are particularly useful for the generation of polynucleotide probes/primers that may be used to quantitate the amount 15 of HGV nucleic acid sequences produced in a cell culture system. Such quantification can be accomplished in a number of ways. For example, probes labelled with reporter molecules can be used in standard dot-blot hybridizations or competition assays of labelled probes with 20 infected cell nucleic acids. Further, there are a number of methods using the polymerase chain reaction to quantitate target nucleic acid levels in a sample 25 (Osikowicz, et al.).

Protective antibodies can also be identified using the cell culture and animal model systems described above. For example, polyclonal or monoclonal antibodies are generated against the antigens of the present invention. 30 These antibodies are then used to pre-treat an infectious HGV-containing inoculum (e.g., serum) before infection of cell cultures or animals. The ability of a single antibody or mixtures of antibodies to protect the cell culture or animal from infection is evaluated. For 35 example, in cell culture and animals the absence of viral antigen and/or nucleic acid production serves as a screen. Further in animals, the absence of HGV hepatitis disease

symptoms, e.g., elevated ALT values, is also indicative of the presence of protective antibodies.

Alternatively, convalescent sera can be screened for the presence of protective antibodies and then these sera used to identify HGV hepatitis associated agent antigens that bind with the antibodies. The identified HGV antigen is then recombinantly or synthetically produced. The ability of the antigen to generate protective antibodies is tested as above.

After initial screening, the antigen or antigens identified as capable of generating protective antibodies, either singly or in combination, can be used as a vaccine to inoculate test animals. The animals are then challenged with infectious HGV. Protection from infection indicates the ability of the animals to generate antibodies that protect them from infection (humoral immunity). Further, use of the animal models allows identification of antigens that activate cellular immunity.

20

#### G. VACCINES AND THE GENERATION OF PROTECTIVE IMMUNITY.

Vaccines can be prepared from one or more of the immunogenic polypeptides identified by the method of the present invention. Genomic organization similarities between the isolated sequences from HGV and other known viral proteins may provide information concerning the polypeptides that are likely to be candidates for effective vaccines. In addition, a number of computer programs can be used for to identify likely regions of isolated sequences that encode protein antigenic determinant regions (for example, Hopp, et al.; "ANTIGEN," Intelligenetics, Mountain View CA).

Vaccines containing immunogenic polypeptides as active ingredients are typically prepared as injectables either as solutions or suspensions. Further, the immunogenic polypeptides may be prepared in a solid or lyophilized state that is suitable for resuspension, prior to

injection, in an aqueous form. The immunogenic polypeptides may also be emulsified or encapsulated in liposomes. The polypeptides are frequently mixed with pharmaceutically acceptable excipients that are compatible with the polypeptides. Such excipients include, but are not limited to, the following and combinations of the following: saline, water, sugars (such as dextrose and sorbitol), glycerol, alcohols (such as ethanol [EtOH]), and others known in the art. Further, vaccine preparations may contain minor amounts of other auxiliary substances such as wetting agents, emulsifying agents (e.g., detergents), and pH buffering agents. In addition, a number of adjuvants are available which may enhance the effectiveness of vaccine preparations. Examples of such adjuvants include, but are not limited to, the following: the group of related compounds including N-acetyl-muranyl-L-threonyl-D-isoglutamine and N-acetyl-nor-muranyl-L-alanyl-D-isoglutamine, and aluminum hydroxide.

The immunogenic polypeptides used in the vaccines of the present invention may be recombinant, synthetic or isolated from, for example, attenuated HGV particles. The polypeptides are commonly formulated into vaccines in neutral or salt forms. Pharmaceutically acceptable organic and inorganic salts are well known in the art.

HGV hepatitis associated agent vaccines are parenterally administered, typically by subcutaneous or intramuscular injection. Other possible formulations include oral and suppository formulations. Oral formulations commonly employ excipients (e.g., pharmaceutical grade sugars, saccharine, cellulose, and the like) and usually contain within 10-98% immunogenic polypeptide. Oral compositions take the form of pills, capsules, tablets, solutions, suspensions, powders, etc., and may be formulated to allow sustained or long-term release. Suppository formulations use traditional binders and carriers and typically contain between 0.1% and 10% of the immunogenic polypeptide.

In view of the above information, multivalent vaccines against HGV hepatitis associated agents can be generated which are composed of one or more structural or non-structural viral-agent polypeptides. These vaccines  
5 can contain, for example, recombinant expressed HGV polypeptides, polypeptides isolated from HGV virions, synthetic polypeptides or assembled epitopes in the form of mosaic polypeptides. In addition, it may be possible to prepare vaccines, which confer protection against HGV  
10 hepatitis infection through the use of inactivated HGV. Such inactivation might be achieved by preparation of viral lysates followed by treatment of the lysates with appropriate organic solvents, detergents or formalin.

Vaccines may also be prepared from attenuated HGV  
15 strains. Such attenuated HGV may be obtained utilizing the above described cell culture and/or animal model systems. Typically, attenuated strains are isolated after multiple passages *in vitro* or *in vivo*. Detection of attenuated strains is accomplished by methods known in the  
20 art. One method for detecting attenuated HGV is the use of antibody probes against HGV antigens, sequence-specific hybridization probes, or amplification with sequence-specific primers for infected animals or assay of HGV-infected *in vitro* cultures.

25 Alternatively, or in addition to the above methods, attenuated HGV strains may be constructed based on the genomic information that can be obtained from the information presented in the present specification. Typically, a region of the infectious agent genome that encodes, for  
30 example, a polypeptide that is related to viral pathogenesis can be deleted. The deletion should not interfere with viral replication. Further, the recombinant attenuated HGV construct allows the expression of an epitope or epitopes that are capable of giving rise to  
35 protective immune responses against the HGV. The desired immune response may include both humeral and cellular immunity. The genome of the attenuated HGV is then used

to transform cells and the cells grown under conditions that allow viral replication. Such attenuated strains are useful not only as vaccines, but also as production sources of viral antigens and/or HGV particles.

5        Hybrid particle immunogens that contain HGV epitopes can also be generated. The immunogenicity of HGV epitopes may be enhanced by expressing the epitope in eucaryotic systems (e.g., mammalian or yeast systems) where the epitope is fused or assembled with known particle forming 10 proteins. One such protein is the hepatitis B surface antigen. Recombinant constructs where the HGV epitope is directly linked to coding sequence for the particle forming protein will produce hybrid proteins that are immunogenic with respect to the HGV epitope and the 15 particle forming protein. Alternatively, selected portions of the particle-forming protein coding sequence, which are not involved in particle formation, may be replaced with coding sequences corresponding to HGV epitopes. For example, regions of specific immunoreactivity 20 to the particle-forming protein can be replaced by HGV epitope sequences.

The hepatitis B surface antigen has been shown to be expressed and assembled into particles in the yeast *Saccharomyces cerevisiae* and in mammalian cells (Valenzuela, 25 et al., 1982 and 1984; Michelle, et al.). These particles have been shown to have enhanced immunoreactivity. Formation of these particles using hybrid proteins, i.e., recombinant constructs with heterologous viral sequences, has been previously disclosed (EPO 175,261, published 26 30 March 1986). Such hybrid particles containing HGV epitopes may also be useful in vaccine applications.

The vaccines of the present invention are administered in dosages compatible with the method of formulation, and in such amounts that will be pharmacologically 35 effective for prophylactic or therapeutic treatments. The quantity of immunogen administered depends on the subject being treated, the capacity of the treatment subject's

immune system for generation of protective immune response, and the desired level of protection.

HGV vaccines of the present invention can be administered in single or multiple doses. Dosage regimens are 5 also determined relative to the treatment subject's needs and tolerances. In addition to the HGV immunogenic polypeptides, vaccine formulations may be administered in conjunction with other immunoregulatory agents.

In an additional approach to HGV vaccination, DNA 10 constructs encoding HGV proteins under appropriate regulatory control are introduced directly into mammalian tissue, *in vivo*. Introduction of such constructs produces "genetic immunization". Similar DNA constructs have been shown to be taken up by cells and the encoded proteins 15 expressed (Wolf, et al.; Ascadi, et al.). Injected DNA does not appear to integrate into host cells chromatin or replicate. This expression gives rise to substantial humoral and cellular immune responses, including protection from *in vivo* viral challenge in animal systems 20 (Wang, et al., 1993; Ulmer, et al.). In one embodiment, the DNA construct is injected into skeletal muscle following pre-treatment with local anesthetics, such as, bupivacaine hydrochloride with methylparaben in isotonic saline, to facilitate cellular DNA uptake. The injected 25 DNA constructs are taken up by muscle cells and the encoded proteins expressed.

Compared to vaccination with soluble viral subunit proteins, genetic immunization has the advantage of authentic *in vivo* expression of the viral proteins. These 30 viral proteins are expressed in association with host cell histocompatibility antigens, and other proteins, as would occur with natural viral infection. This type of immunization is capable of inducing both humoral and cellular immune responses, in contrast to many soluble 35 subunit protein vaccines. Accordingly, this type of immunization retains many of the beneficial features of

live attenuated vaccines, without the use of infectious agents for vaccination and attendant safety concerns.

Direct injection of plasmid or other DNA constructs encoding the desired vaccine antigens into *in vivo* tissues 5 is one delivery means. Other means of delivery of the DNA constructs can be employed as well. These include a variety of lipid-based approaches in which the DNA is packaged using liposomes, cationic lipid reagents or cytofectins (such as, lipofectin). These approaches 10 facilitate *in vivo* uptake and expression, as summarized by Felgner and Rhodes (1991). Various modifications to these basic approaches include the following: incorporation of peptides, or other moieties, to facilitate (i) targeting to particular cells, (ii) the intracellular disposition of 15 the DNA construct following uptake, or (iii) to facilitate expression. Alternatively, the sequences encoding the desired vaccine antigens may be inserted into a suitable retroviral vector. The resulting recombinant retroviral vector inoculated into the subject for *in vivo* expression 20 of the vaccine antigen. The antigen then induces the immune responses. As noted above, this approach has been shown to induce both humoral and cellular immunity to viral antigens (Irwin, et al.).

Further, the HGV vaccines of the present invention 25 may be administered in combination with other vaccine agents, for example, with other hepatitis vaccines.

#### H. SYNTHETIC PEPTIDES.

When the coding sequences of HGV polypeptide antigens 30 are determined synthetic peptides can be generated which correspond to these polypeptides. Synthetic peptides can be commercially synthesized or prepared using standard methods and apparatus in the art (Applied Biosystems, Foster City CA).

35 Alternatively, oligonucleotide sequences encoding peptides can be either synthesized directly by standard methods of oligonucleotide synthesis, or, in the case of

large coding sequences, synthesized by a series of cloning steps involving a tandem array of multiple oligonucleotide fragments corresponding to the coding sequence (Crea; Yoshio et al.; Eaton et al.). Oligonucleotide coding sequences can be expressed by standard recombinant procedures (Maniatis et al.; Ausubel et al.).

IV. CHARACTERIZATION OF THE VIRAL GENOME.

As shown in Example 4, the HGV genome appears to be an RNA molecule and has the closest sequence similarity to viral sequences that are categorized in the Flaviviridae family of viruses. This family includes the Flaviviruses, Pestiviruses and an unclassified Genus made up of one member, Hepatitis C virus. The HGV virus does not have significant global (i.e., over the length of the virus) sequence identity with other established members of the Flaviviridae -- with the exception of the protein motifs discussed below.

In general members of the Flaviviridae are enveloped viruses that have densities in sucrose gradients between 1.1 and 1.23 g/ml and are sensitive to heat, organic solvents and detergents. As shown in Example 5, HGV has density characteristics similar to an enveloped Flaviviridae virus (HCV). The integrity of the HGV virion also appears to be sensitive to organic solvents (Example 5).

Flaviviridae virions contain a single molecule of linear single-stranded (ss) RNA which also serves as the only mRNA that codes for the viral proteins. The ssRNA molecule is typically between the size of 9 and 12 kilobases long.

Viral proteins are derived from one polyprotein precursor that is subsequently processed to the mature viral proteins. Most members of the Flaviviridae do not contain poly(A) tails at their 3' ends. Virion are about 15-20% lipid by weight.

Members in the Flaviviridae family have a core protein and two or three membrane-associated proteins. The analogous structural proteins of members in the three genera Flavivirus family show little similarity to one another at the sequence level. The nonstructural proteins contain conserved motifs for RNA dependent RNA polymerase (RDRP), helicase, and a serine protease. These short blocks of conserved amino acids or motifs can be detected using computer algorithms known in the art such as "MACAW" (Schuler, et al.). These motifs are presumably related to constraints imposed by substrates processed by these proteins (Koonin and Dolja). The order of these motifs is conserved in all members of the Flaviviridae family. The genome of HGV contains at least the protein motifs found in the RNA dependent RNA polymerase (RDRP) of members of the Flaviviridae family (see Figure 5, "GDD" sequence).

Members of the Flaviviridae family are known to replicate in a wide variety of animals ranging from (i) hematophagous arthropod vectors (ticks and mosquitoes), where they do not cause disease, to (ii) a large range of vertebrate hosts (humans, primates, other mammals, marsupials, and birds). Over 30 members of the Flaviviridae family cause diseases in man, ranging from febrile illness, or rash, to potentially fatal diseases such as hemorrhagic fever, encephalitis, or hepatitis. At least 10 members of the Flaviviridae family cause severe and economically important diseases in domestic animals.

30 V. DETECTION OF ANTIGENS CODED BY SHORT REVERSE READING FRAMES COINCIDENT WITH KNOWN READING FRAMES

The present invention provides antigens useful for the determination of whether a test subject (e.g., human patient or animal) has been infected with a virus having an RNA genome, and a method for identifying such antigens. RNA viruses include, but are not limited to, the following families: Picornaviridae, Caliciviridae, Reoviridae, Birnaviridae, Togaviridae, Flaviviridae, Orthomyxoviridae,

Paramyxoviridae, Rhabdoviridae, Filoviridae, Coronaviridae, Bunyaviridae, Retroviridae, and Arenaviridae. These families include single- and double-stranded RNA genomes, segmented and non-segmented genomes.

- 5 In a preferred embodiment, the method of the present invention is applied to RNA viruses having single-strand genomes.

The method of the present invention teaches the expression and subsequent induction of antibodies to a 10 protein or proteins coded by "reverse reading frames" of RNA viruses. "Reverse reading frames" are defined as open reading frames that are transcribed and translated in the opposite direction to the major known reading frames for the virus, i.e., identifiable viral proteins.

15 Identification of reverse-reading frame encoded antigens can be accomplished as follows. Coding regions of viral polynucleotides are examined to determine the coding regions corresponding to coding sequences for identifiable viral proteins. Such identifiable viral 20 proteins include, for example, typical viral structural (e.g., capsid) and non-structural (e.g., RNA dependent RNA polymerase, reverse transcriptase, and proteases) proteins. A further example of such identifiable viral proteins includes the polyprotein of members of Flavi- 25 viridae.

The complement (i.e., the reverse frame) of the polynucleotide strand encoding identifiable viral protein(s) is evaluated for open reading frames using the following method. First, conserved open frames are identified among the complement strands of variants of a 30 selected virus. Typically, variants are chosen that show low global sequence identity conservation relative to each other. A program such as DM.EXE (MS-DOS program from David Mount and Bruce Conrad, University of Arizona, 35 Tucson, AZ) or alternatively the PC/GENE suite of programs (Intelligenetics, Mountain View, CA) facilitates the

identification of open reading frames in the reverse frame.

Reverse open reading frames that are conserved between, for example, two variants are then examined in 5 other isolates. Reverse open reading frames that are conserved in a number of variants of a virus (e.g., among many HCV variants) are candidates for reverse frame antigens. As longer reverse open reading frames are more difficult to conserve, the longest frames should be examined first.

In general, the starting codons of the frames are conserved but minor variations of the terminations and length can be accepted. Frames can be as short as about 12 amino acids, but preferably the reading frame is at 15 least about 30 amino acids in length, and even more preferably at least about 30 to 100 amino acids in length.

Although it is preferred to compare variants for conserved reverse open reading frames, it is also within the scope of the invention to select any reverse open 20 reading frame and screen the encoded protein, as described below, for antigenic activity.

After identification of reverse-frame coding sequences, the polypeptide encoded by the sequence is produced, for example, recombinantly or synthetically 25 (e.g., solid phase chemical synthesis). In one embodiment, recombinant proteins coded by the reverse open reading frames are expressed in *E. coli* expression systems. The antigens are screened against sera known to be specifically immunoreactive with viral antigens from the 30 virus whose genome is being evaluated. For example, the antigens are used to detect antibodies in humans or animals infected with RNA viruses. Specific examples are given below for HGV and HCV.

The diagnostic utility of reverse-frame antigens 35 identified by this method are evaluated using immunological screening of panels of sera known or suspected to be infected with the viral agent from which the reverse frame

antigens were derived. Exemplary embodiments of antigen selection using this method, and use of such antigens in diagnostic assays, are described below.

5       A. Detection of Viral Antibodies

The method of the present invention includes detection of viral antibodies based on the detection of an antigen coded by the reverse reading frame from the expected major coding open frame. In one embodiment of the 10 present invention, a reverse reading frame antigen was identified for the RNA virus HGV: the antigen encoded by the 470-20-1 clone was detected with antibodies from several N-(ABCDE) hepatitis sera, including PNF 2161. The sequence of the 470-20-1 clone was extended by Anchored 15 PCR cloning (Example 6).

Analysis of the regions surrounding the original clone 470-20-1 open reading frame revealed an extended open reading frame of approximately 161 amino acids (SEQ ID NO:28). Analysis of the opposite strand to the protein 20 coding strand of 470-20-1 revealed that it consisted of a completely open reading frame for a polyprotein sequence (Figure 1). Similarity analysis of the polyprotein detected sequence similarity to members of the Flaviviridae family (see Section IV).

25       All members of Flaviviridae code for their known viral proteins using a long open reading frame to produce a polyprotein that is subsequently processed to the individual viral proteins. The sequence similarity of HGV to Flaviviridae is seen in the long, open, reverse-reading 30 frame relative to the coding sequences for the 420-20-1 antigen -- implying that the 470-20-1 antigen is actually coded in the opposite direction from the expected major coding region. Yet, the 470-20-1 antigen has been useful to detect infection of sera by HGV (Example 9).

35       Further reverse-frame HGV antigens have been identified as follows. Three distinct immunogenic regions were isolated from three different HGV-epitope libraries. All

three epitopic regions are encoded by the negative strand (i.e., the opposite strand relative to the strand encoding the polyprotein) of the HGV virus. The antigenic regions encoded by the negative strand are all contained within 5 relatively short and separate open reading frames (ORFs). The three libraries constructed for screening are described below.

The first immunogenic region is defined by a single clone K1-2-3a (SEQ ID NO:111; SEQ ID NO:112). K1-2-3 was 10 isolated from a library designated NS3 which was generated by polymerase chain reaction amplification from PNF 2161 serum nucleic acids using the primer set 470ep-f9 (SEQ ID NO:98) and 470ep-R9 (SEQ ID NO:99). These primers amplify a fragment of HGV from the NS3 region. Fragment F9/R9 was 15 amplified from 1 µl of PNF 2161 SISPA amplified DNA. Amplifications were for 30 cycles for 1 minute at 94°C, 2 minutes at 52°C and 3 minutes at 72°C. The expected 777 nucleotide product was gel purified.

The primers were also used for amplification of the 20 same fragment from a larger clone that was also obtained from PNF 2161 serum nucleic acids. The two purified DNA fragments were combined and partially digested with DNase I. The partially digested sample (designated the F9/R9 library) was ligated to KL1 SISPA linkers and digested 25 with EcoRI. The F9/R9 DNA was ligated into lambda gt11 and packaged.

The clone K1-2-3a was isolated by screening of the library expressing the F9/R9 fragment. Ten plates at 30,000 plaques/plate were screened with PNF 2161 plasma 30 diluted 1/100 in AIB. Twenty two first round positive plaques were identified. Clone K1-2-3a was purified from one of these plaques and was repeatedly immunoreactive against PNF 2161 sera.

Sequencing of the K1-2-3a clone (SEQ ID NO:111; SEQ 35 ID NO:112) indicated that it expresses a 44 amino acid insert. Analysis of the position of the K1-2-3a sequence with respect to the sequence of the negative strand of HGV

indicated K1-2-3 is contained within a 100 amino acid ORF that is located in the negative strand of the NS3 gene of HGV. This ORF contains 1 methionine. The total size of ORF from the methionine to the termination codon is 51 5 amino acids. This methionine residue is also contained within the K1-2-3 sequence at position 4.

The next reverse-frame immunogenic region was designated the K3 region. The K3 series of clones was isolated from a library designated NS2. The library was generated 10 using the primers given in Table 1 and SISPA amplified PNF 2162 DNA as template.

Table 1

15	Fragments	nt	
	9E3-REV (SEQ ID NO:100) E39-94PR (SEQ ID NO:101)	59 2	aa 358 (of 389) of E2 to aa 166 of NS-2
	GEP-F12 (SEQ ID NO:102) GEP-R12 (SEQ ID NO:106)	66 3	aa 144 (of 313) of NS-2 to aa 51 of NS-3
20	GEP-F14 (SEQ ID NO:103) GEP-R13 (SEQ ID NO:107)	71 5	aa 357 - 594 of NS-3
	470epF8 (SEQ ID NO:97) GEP-R14 (SEQ ID NO:108)	64 8	aa 716 - 847 of NS-5 (716 to end)

25

All amplifications were for 35 cycles of 94°C/1 minute, 48°C/2 minutes, and 73°C/3 minutes. All amplifications yielded at least a fragment of the expected size. The amplified products were mixed and in an approximately 30 1:1:1:1 ratio and partially digested with DNaseI. As above, the digestion products were ligated to KL1 SISPA linkers, amplified and EcoRI digested. The digested fragments were ligated into lambda gt11. The ligation reactions were packaged.

35 The packaged ligation products were plated. Screening of this epitope library with PNF 2161 serum resulted in the isolation of 35 putatively immunoreactive plaques. Of the 35 positive areas, 22 were reportedly immunoreac-

tive with PNF 2161 serum. Twelve of the positive plaques were purified, re-screened and sequenced.

Eight of the 12 clones contained essentially the same insert (not counting repeated sequences and linkers).

- 5 These clones are K3-8-5A (SEQ ID NO:131; SEQ ID NO:132), K3-10-1D (SEQ ID NO:113; SEQ ID NO:114), K3-8-4C (SEQ ID NO:129; SEQ ID NO:130), K3-8-7C (SEQ ID NO:135; SEQ ID NO:136), K3-14-3A (SEQ ID NO:119; SEQ ID NO:120), K3-14-6A (SEQ ID NO:123; SEQ ID NO:124), K3-14-2A (SEQ ID NO:117; SEQ ID NO:118), and K3-14-5A (SEQ ID NO:121; SEQ ID NO:122). One of the 12 was the same as these 8 clones except for a 3 nt insertion (K3-17-1A; SEQ ID NO:125, SEQ ID NO:126).

- 15 One of the 12 clones was a unique chimera (K3-8-3A; SEQ ID NO:127, SEQ ID NO:128). Two of the 12 clones were unique long clones (K3-11-1A -- SEQ ID NO:115, SEQ ID NO:116; and K3-8-6A -- SEQ ID NO:133, SEQ ID NO:134).

- 20 All of the K3 clones express the negative strand of HGV (i.e., relative to the coding strand for the polyprotein). All of the K3 clones have completely open reading frames through their entire inserts. An alignment of these clones is presented as Figures 11A, 11B and 11C.

- The K3 clones are contained with the PCR fragment derived from amplification with the 9e3-rev (SEQ ID NO:100) and E39-94pr (SEQ ID NO:101) primers. This fragment contains the COOH terminal 31 amino acids of HGV E2 gene and the amino terminal 166 amino acids of HGV, NS2 gene.

- 30 All of the K3 clones contain a frame shift relative to the consensus sequence of the reverse strand of HGV: 11 of the 12 clones are missing 1 C residue; and the 12th clone (K3-17-1) contains 3 additional C residues.

- 35 The 5' end of all of the K3 clones is contained within a 171 amino acid ORF of the negative strand. This ORF contains a methionine at position 23, such that the greatest possible length of the methionine to termination codon open reading frame is 149 amino acids (approximately

18 kd). All of the K3 clones (except K3-8-6) have their 5' terminal defined by the PCR primer E39-94pr (SEQ ID NO:101), which corresponds to amino acid 87 of the 171 acid ORF. All of the clones continue in this ORF until 5 the occurrence of the frame shift at amino acid 140. At this point, all clones frame shift into the 8th amino acid of a new ORF (Figure 11B). The clones all then express the sequence SEQ ID NO:149.

Then the reading frames of all the clones, except K3-10 8-6 and K3-11-1, shift to an 8 nucleotide sequence of unknown origin (coding the amino acids QHS) then into the sequence of the reverse primer 9e3-rev (SEQ ID NO:100) which expresses the amino acids SEQ ID NO:148 (Figure 11C). SEQ ID NO:148 is in the same frame as the common 15 sequence SEQ ID NO:147 at amino acid 277 of the long combined frame (amino acid 144 of the 2nd frame).

The 2 clones K3-11-1 and K3-8-6 are co-linear with the new frames until their inserts end at amino acids 192 and 259.

20 In summary, this group of clones contains multiple disparately located sequences, whose final contribution to the observed immunoreactivity is being determined. Primers for the subcloning of various permutations of the amino acid sequences from the K3 region have been 25 designed. Subfragments of the K3 region will be cloned into the expression vector pGEX-HIS-B. Preliminary data confirms that 2 of these sequences are highly immunoreactive with PNF 2161 sera when expressed as a fusion protein with sj26.

30 The last negative strand immunogenic region is defined by the clones Y10-13-1 (SEQ ID NO:137; SEQ ID NO:138) and Y10-13-2 (SEQ ID NO:139; SEQ ID NO:140). These clones were derived from the envelope protein coding region. The env library was generated by PCR amplification of 1  $\mu$ l of PNF 2161 SISPA-amplified material 35 using the primers presented in Table 2.

Table 2

Fragments	nt	
5 GEP-F15 (SEQ ID NO:104) GEP-R15 (SEQ ID NO:109)	52 5	-182 amino acid of the COOH $\frac{1}{2}$ of E2
GEP-F17 (SEQ ID NO:110) GEP-R16 (SEQ ID NO:105)	76 5	the COOH term of E1 through ~ aa 220 of E2

10

PCR amplification was for 35 cycles of 94°C/1 minute, 52°C/1.5 minutes, 72°C/3 minutes. The amplified products were purified, partially digested with DNaseI, and ligated to KL1 linkers. The ligated KL1 DNAs were amplified, 15 digested with EcoRI and ligated into lambda gt11. This library was screened with the HGV positive sera R34587: 150,000 recombinant phage were screened. From this screening positive areas were isolated, plaque purified and re-screened. Three plaques were identified that were 20 repeatedly reactive with R34587 sera. Two of these plaques, Y10-13-1 and Y10-13-2, were sequenced.

The clones Y10-13-1 and Y10-13-2 are contained within the PCR fragment defined by GEP-F17 and GEP-R16. The inserts of both clones represent continuous open reading 25 frames. They are contained within a 139 amino acid ORF of the negative strand. This ORF has a methionine present at amino acid 22 (where the longest open reading frame is 117 amino acids, methionine to termination codon). Both clones start downstream of the methionine (Y10-13-1 = 30 amino acids 39-116 of the ORF; Y10-13-2 = amino acids 57-116 of the ORF). The epitopes in all of the above clones will be mapped.

Further reverse-frame HGV antigens can be identified using the above-described methods and a selected HGV 35 polynucleotide (e.g., SEQ ID NO:14 or SEQ ID NO:156, Example 13).

B. Reverse-Reading Frame Encoded Antigens in Other RNA Viruses.

The virus HCV is a member of the Flaviviridae family. Three members of the HCV group of viruses were analyzed 5 for conserved, reverse open reading frames: (accession numbers/viral designation, Genbank Ver. 83, Intelligenetics, Mountain View, CA) M58335/HPCHUMR; D90208/-HPCJCG; and M62321/HPCPLYPRE. Two exemplary reverse open reading frames were identified that were conserved between 10 the three members. Each of these open reading frames start with a methionine codon and end at a termination codon. Figure 10 shows a schematic of the inverse sequence of the HCV genome based on the 9401 base pair sequences obtained from isolate HPCPLYPRE. The open boxes 15 in Figure 10 show several exemplary open reading frames; inverse ORF1 and inverse ORF2 represent the position of the two conserved open reading frames. The coordinates for these open reading frames are presented in Table 3.

20

Table 3

Virus	ORF	Start	End	ORF Size	SEQ ID NO:
M62321	1r	2876	3259	128	141
	2r	3404	3835	144	142
M58335	1r	2900	3199	107	143
	2r	3533	3934	134	144
D90208	1r	2900	3220	100	145
	2r	3533	3935	134	146

25

Coordinates are expressed as number of base pairs from the 3' end of the positive strand of the virus.

30

The present invention provides a novel method to determine whether a test subject has been infected with a virus. Experiments performed in support of the present 35 invention suggest the expression and subsequent induction of antibodies to a polypeptide or polypeptides coded by

reverse frames in the opposite direction of the major known reading frames of RNA viruses. This phenomena forms the basis of a diagnostic assay based on detection of antibodies directed against polypeptide antigens coded for 5 by the reverse frame of RNA viruses.

The reverse-frame antigens of the present invention can be utilized in the applications exemplified herein for HGV embodiments, for example, vaccine, antibodies, methods and diagnostics.

10

## VI. Utility

### A. IMMUNOASSAYS FOR HGV.

One utility for the antigens obtained by the methods of the present invention is their use as diagnostic reagents for the detection of antibodies present in the sera 15 of test subjects infected with HGV hepatitis virus, thereby indicating infection in the subject; for example, 470-20-1 antigen, antigens encoded by SEQ ID NO:14 or its complement, and antigens encoded by portions of either 20 strand of the complete viral sequence. The antigens of the present invention can be used singly, or in combination with each other, in order to detect HGV. The antigens of the present invention may also be coupled with diagnostic assays for other hepatitis agents such as HAV, 25 HBV, HCV, and HEV.

In one diagnostic configuration, test serum is reacted with a solid phase reagent having a surface-bound antigen obtained by the methods of the present invention, e.g., the 470-20-1 antigen. After binding with anti-HGV 30 antibody to the reagent and removing unbound serum components by washing, the reagent is reacted with reporter-labelled anti-human antibody to bind reporter to the reagent in proportion to the amount of bound anti-HGV antibody on the solid support. The reagent is again 35 washed to remove unbound labelled antibody, and the amount of reporter associated with the reagent is determined. Typically, the reporter is an enzyme which is detected by

incubating the solid phase in the presence of a suitable fluorometric or colorimetric substrate (Sigma, St. Louis, MO).

- The solid surface reagent in the above assay is
- 5 prepared by known techniques for attaching protein material to solid support material, such as polymeric beads, dip sticks, 96-well plate or filter material. These attachment methods generally include non-specific adsorption of the protein to the support or covalent attachment
- 10 of the protein, typically through a free amine group, to a chemically reactive group on the solid support, such as an activated carboxyl, hydroxyl, or aldehyde group.
- Alternatively, streptavidin coated plates can be used in conjunction with biotinylated antigen(s).
- 15 Also forming part of the invention is an assay system or kit for carrying out this diagnostic method. The kit generally includes a support with surface-bound recombinant HGV antigen (e.g., the 470-20-1 antigen, as above), and a reporter-labelled anti-human antibody for
- 20 detecting surface-bound anti-HGV antigen antibody.

In a second diagnostic configuration, known as a homogeneous assay, antibody binding to a solid support produces some change in the reaction medium which can be directly detected in the medium. Known general types of

25 homogeneous assays proposed heretofore include (a) spin-labelled reporters, where antibody binding to the antigen is detected by a change in reported mobility (broadening of the spin splitting peaks), (b) fluorescent reporters, where binding is detected by a change in fluorescence

30 efficiency or polarization, (c) enzyme reporters, where antibody binding causes enzyme/substrate interactions, and (d) liposome-bound reporters, where binding leads to liposome lysis and release of encapsulated reporter. The adaptation of these methods to the protein antigen of the

35 present invention follows conventional methods for preparing homogeneous assay reagents.

In each of the assays described above, the assay method involves reacting the serum from a test individual with the protein antigen and examining the antigen for the presence of bound antibody. The examining may involve

5 attaching a labelled anti-human antibody to the antibody being examined (for example from acute, chronic or convalescent phase) and measuring the amount of reporter bound to the solid support, as in the first method, or may involve observing the effect of antibody binding on a

10 homogeneous assay reagent, as in the second method.

A third diagnostic configuration involves use of HGV antibodies capable of detecting HGV-specific antigens. The HGV antigens may be detected, for example, using an antigen capture assay where HGV antigens present in candidate serum samples are reacted with a HGV specific monoclonal or polyclonal antibody. The antibody is bound to a solid substrate and the antigen is then detected by a second, different labelled anti-HGV antibody. Antibodies can be prepared, utilizing the peptides of the present

15 invention, by standard methods. Further, substantially isolated antibodies (essentially free of serum proteins which may affect reactivity) can be generated (e.g., affinity purification (Harlow et al.)).

25        B. HYBRIDIZATION ASSAYS FOR HGV.

One utility for the nucleic acid sequences obtained by the methods of the present invention is their use as diagnostic agents for HGV sequences present in sera, thereby indicating infection in the individual. Primers

30 and/or probes derived from the coding sequences of the present invention, in particular, Clone 470-20-1 and SEQ ID NO:14, can be used singly, or in combination with each other, in order to detect HGV.

In one diagnostic configuration, test serum is reacted under PCR or RT-PCR conditions using primers derived from, for example, 470-20-1 sequences. The presence of HGV, in the serum used in the amplification reaction, can

be detected by specific amplification of the sequences targeted by the primers. Example 4 describes the use of polymerase chain amplification reactions, employing primers derived from the clones of the present invention, 5 to screen different source material. The results of these amplification reactions demonstrate the ability of primers derived from the clones of the present invention (for example, 470-20-1), to detect homologous sequences by amplification reactions employing a variety of different 10 source templates. The amplification reactions in Example 4 included use of nucleic acids obtained directly from sera as template material.

Alternatively, probes can be derived from the HGV sequences of the present invention. These probes can then 15 be labelled and used as hybridization probes against nucleic acids obtained from test serum or tissue samples. The probes can be labelled using a variety of reporter molecules and detected accordingly: for example, radioactive isotopic labelling and chemiluminescent detection 20 reporter systems (Tropix, Bedford, Mass.).

Target amplification methods, embodied by the polymerase chain reaction, the self-sustained sequence replication technique ["3SR," (Guatelli, et al.; Gingeras, et al., 1990) also known as "NASBA" (VanGemen, et al.)], the 25 ligase chain reaction (Barany), strand-displacement amplification ["SDA," (Walker)], and other techniques, multiply the number of copies of the target sequence. Signal amplification techniques, exemplified by branched-chain DNA probes (Horn and Urdea; Urdea; Urdea, et al.) and the 30 Q-beta replicase method (Cahill, et al.; Lomell, et al.), first bind a specific molecular probe, then replicate all of or part of this probe or in some other manner amplify the probe signal.

For the detection of the specific nucleic acid sequences disclosed in the present invention or contiguous 35 sequences in the same or a similar (related) viral genome, amplification and detection methodologies may be employed,

as alternatives to amplification by the PCR. A number of such techniques are known to the field of nucleic acid diagnostics (The 1992 San Diego Conference: Genetic Recognition, *Clin. Chem.* 39(4):705 (1993)).

5

1. SELF-SUSTAINED SEQUENCE REPLICATION.

The Self-Sustained Sequence Replication (3SR) technique results in amplification to a similar magnitude as PCR, but isothermally. Rather than thermal cycle-driven 10 PCR, the 3SR operates as a concerted three-enzyme reaction of a) cDNA synthesis by reverse transcriptase, b) RNA strand degradation by RNase H, and c) RNA transcription by T7 RNA polymerase.

As the entire reaction sequence occurs isothermally 15 (typically at 42°C.), expensive temperature-cycling instrumentation is not required. In the absence of duplex denaturation via heating, organic solvents, or other mechanism, only single-stranded templates (i.e., predominantly RNA) are amplified.

20 Suitable primers for use in 3SR amplification can be selected from the viral sequences of the present invention by those having ordinary skill in the art. For example, for isothermal amplification of viral sequences by the 3SR technique, primer 470-20-1-77F (SEQ ID NO:9) is modified 25 by the addition of the T7 promoter sequence and a preferred T7 transcription initiation site to the 5'-end of the oligonucleotide. This modification results in a suitable 3SR primer T7-470-20-1-77F (SEQ ID NO:9). Primer 470-20-1-211R (SEQ ID NO:10) can be used in these 30 reactions either without modification or T7 promoter.

RNA extracted from PNF 2161 is incubated with AMV reverse transcriptase (30 U), RNase H (3 U), T7 RNA polymerase (100 U), in 100 ul reactions containing 20 mM Tris-HCl, pH 8.1 (at room temperature), 15 mM MgCl<sub>2</sub>, 10 mM KCl, 35 2 mM spermidine HCl, 5 mM dithiothreitol (DTT), 1 mM each of dATP, dCTP, dGTP, and TTP, 7 mM each of ATP, CTP, GTP,

and UTP, and 0.15 uM each primer. Amplification takes place during incubation at 42°C. for 1-2 h.

Initially, primer T7-470-20-1-77F anneals to the target RNA, and is extended by AMV reverse transcriptase 5 to form cDNA complementary to the starting RNA strand. Following degradation of the RNA strand by RNase H, reverse transcriptase catalyzes the synthesis of the second strand DNA, resulting in a double-stranded template containing the (double-stranded) T7 promoter sequence. RNA 10 transcription results in production of single-stranded RNA. This RNA then serves to re-enter the cycle for additional rounds of amplification, finally resulting in a pool of high-concentration product RNA. The product is predominantly single-stranded RNA of the same strand as 15 the primer containing the T7 promoter (T7-470-20-1-77F), with much smaller amounts of cDNA.

Alternatively, the other primer (470-20-1-211R) may contain the T7 promoter, or both primers may contain the promoter, resulting in production of both strands of RNA 20 as products of the reaction. Products of the 3SR reaction may be detected, characterized, or quantitated by standard techniques for the analysis of RNA (e.g., Northern blots, RNA slot or dot blots, direct gel electrophoresis with RNA-staining dyes). Further, the products may be detected 25 by methods making use of biotin-avidin affinity interactions or specific hybridizations of nucleic acid probes.

In one technique for rapid and specific analysis of 3SR products, solution hybridization of the product to 30 radiolabelled oligonucleotide 470-20-1-152R (SEQ ID NO:21) is followed by non-denaturing polyacrylamide gel electrophoresis. This assay (a gel mobility shift-type assay) results in the detection of specific probe-product hybrid as a slower-moving band than the band corresponding 35 to unhybridized oligonucleotide.

## 2. LIGASE CHAIN REACTION (LCR)

As another example of a detection system, the HGV sequence may form the basis for design of ligase chain reaction (LCR) primers. LCR makes use of the nick-closing activity of DNA ligase to join two immediately adjacent oligonucleotides possessing adjacent 5'-phosphate ("donor" oligo) and 3'-hydroxyl ("acceptor" oligo) termini. The property of DNA ligase to join only fully complementary ends in a template-dependent way, leads to a high degree of specificity, in that ligation will not occur unless the termini to be linked are perfectly matched in sequence to the target strand.

As an alternative to PCR, with some advantages in terms of specificity for discrimination of single base mismatches between primer and target nucleic acid, the LCR may be used to detect or "type" strains of virus possessing homology to HGV sequences. These techniques are suitable for assessing the presence of specific mutations when such base changes are known to confer drug resistance (e.g., Larder and Kemp; Gingeras, et al., 1991).

In the presence of template-complementary donor and acceptor oligonucleotides and oligonucleotides complementary to the donor and acceptor, exponential amplification by LCR is possible. In this embodiment, each round of ligation generates additional template for subsequent rounds, in a cyclic reaction.

For example, primer 470-20-1-211R (SEQ ID NO:10), an adjacent oligonucleotide (B, SEQ ID NO:22) and cognate oligos (211R', SEQ ID NO:23, and B', SEQ ID NO:24), can be used to perform LCR amplification of the sequence of this invention. Reverse transcription is first performed by standard methods to generate cDNA, which is then amplified in reactions containing 0.1-1  $\mu$ M each of the four LCR primers, 20 mM Tris-HCl, pH 8.3 (room temperature), 25 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM dithiothreitol (DTT), 0.5 mM NAD+, 0.01% Triton X-100, and 5 Units of DNA ligase (Ampligase,

Epicentre Technologies, Madison, WI, or other commercial supplier of thermostable DNA ligase), in 25 ul reactions.

Thermal cycling is performed at 94°C. for 1 min. 30 s; 94°C. for 1 min., 65°C. for 2 min., repeated for 25-40 cycles. Specificity of product synthesis depends on primer-template match at the 3'-terminal position. Products are detected by polyacrylamide gel electrophoresis, followed by ethidium bromide staining; alternatively, one of the acceptor oligos (211R' or B) is 5'-radiolabelled for visualization by autoradiography following gel electrophoresis.

Alternatively, a donor oligo is 3'-end-labelled with a specific bindable moiety (e.g., biotin), and the acceptor is 5'-labelled with a specific detectable group (e.g., a fluorescent dye), for solid phase capture and detection.

### 3. METHODS FOR ANALYSIS OF AMPLIFIED DNA

Numerous techniques have been described for the analysis of amplified DNA. Several such techniques are advantageous for high-throughput applications, where gel electrophoresis is impractical, for example, rapid and high-resolution HPLC techniques (Katz and Dong). However, in general, methods for infectious disease organism screening using nucleic acid probes involve a separate post-amplification hybridization step in order to assure requisite specificity for pathogen detection.

One such detection embodiment is an affinity-based hybrid capture technique (Holodniy, et al.). In this embodiment the PCR is conducted with one biotinylated primer. Following amplification, the double-stranded product is denatured then hybridized to a peroxidase-labelled probe complementary to the strand having incorporated the biotinylated primer. The hybridized product is then incubated in a buffer which is in contact with an avidin (or streptavidin) coated surface (e.g., membrane filter, microwell, latex or paramagnetic beads).

The mass of coated solid phase which contacts the volume of PCR product to be analyzed by this method must contain sufficient biotin-binding sites to capture essentially all of the free biotinylated primer, as well as the 5 much lower concentration of biotinylated PCR product.

Following three to four washes of the solid phase, bound hybridized product is detected by incubation with o-phenylenediamine in citrate buffer containing hydrogen peroxide.

10 Alternatively, capture may be mediated by probe-coated surfaces, followed by affinity-based detection via the biotinylated primer and an avidin-reporter enzyme conjugate (Whetsell, et al.).

15 4. ADDITIONAL METHODS

Viral sequences of the present invention may also form the basis for a signal amplification approach to detection, using branched-chain DNA probes. Branched-chain probes (Horn and Urdea; Urdea) have been described 20 for detection and quantification of rare RNA and DNA sequences (Urdea, et al.). In this method, an oligonucleotide probe (RNA, DNA, or nucleic acid analogue) is synthesized with a sequence complementary to the target RNA or DNA. The probe also contains a unique branching 25 sequence or sequences not complementary to the target RNA or DNA.

This unique sequence constitutes a target for hybridization of branched secondary detector probes, each of which contains one or more other unique sequences, serving 30 as targets for tertiary probes. At each branch point in the signal amplification pathway, a different unique sequence directs hybridization of secondary, tertiary, etc., detection probes. The last probe in the series typically is linked to an enzyme useful for detection 35 (e.g., alkaline phosphatase). The sequential hybridization of primers eventually results in the buildup

of a highly-branched structure, the arms of which terminate in enzyme-linked probes.

Enzymatic turnover provides a final amplification, and the choice of highly sensitive chemiluminescent substrates (e.g., LumiPhos, Lumigen, Detroit, MI, as a substrate for alkaline phosphatase labels) results in exquisite sensitivity, on the order of 10,000 molecules or less of original target sequence per assay. In such a detection method, amplification depends only on molecular hybridization, rather than enzymatic mechanisms, and is thus far less susceptible to inhibitory substances in clinical specimens than, for example, PCR. Thus, this detection method allows the use of crude techniques for nucleic acid release in test samples, without extensive purification before assay.

Amplification for sensitive detection of the viral sequences of the present invention may also be accomplished by the Q- $\beta$  replicase technique (Cahill, et al.; Lomell, et al.; Pritchard, et al.). In this method, a specific probe is designed to be complementary to the target sequence. This probe is then inserted by standard molecular cloning techniques into the sequence of the replicatable RNA from Q- $\beta$  phage. Insertion into a specific region of the replicon does not prevent replication by Q- $\beta$  replicase.

Following molecular hybridization, and several cycles of washing, the replicase is added and amplification of the probe RNA ensues. "Reversible target capture" is one known technique for reducing the potential background from replication of unhybridized probes (Morrissey, et al.). Amplified replicons are detectable by standard molecular hybridization techniques employing DNA, RNA or nucleic acid analogue probes.

Additional methods for amplification and detection of rare DNA or RNA sequences are known in the literature and preferred to the PCR for some applications in the field of molecular diagnostics. These alternative techniques may

form the basis for detection, characterization (e.g., sequence diversity existing as multiple related strains of the sequence described herein, genotypic changes characteristic of drug resistance), or quantification of 5 the sequence disclosed in the present invention.

Also forming part of the invention are assay systems or kits for carrying out the amplification/hybridization assay methods just described. Such kits generally include either specific primers for use in amplification reactions 10 or hybridization probes.

The following examples illustrate, but in no way are intended to limit the present invention.

15

#### MATERIALS AND METHODS

*E. coli* DNA polymerase I (Klenow fragment) was obtained from Boehringer Mannheim Biochemicals (BMB) (Indianapolis, IN). T4 DNA ligase and T4 DNA polymerase were obtained from New England Biolabs (Beverly, MA); Nitro- 20 cellulose and "NYTRAN" filters were obtained from Schleicher and Schuell (Keene, NH).

Synthetic oligonucleotide linkers and primers were prepared using commercially available automated oligonucleotide synthesizers. Alternatively, custom designed 25 synthetic oligonucleotides may be purchased from commercial suppliers. cDNA synthesis kit and random priming labeling kits were obtained from BMB (Indianapolis, IN) or GIBCO/BRL (Gaithersburg, MD).

Standard molecular biology and cloning techniques 30 were performed essentially as previously described in Ausubel, et al., Sambrook, et al., and Maniatis, et al.

Common manipulations relevant to employing antisera and/or antibodies for screening and detection of immuno-reactive protein antigens were performed essentially as 35 described (Harlow, et al.). Similarly ELISA and Western blot assays for the detection of anti viral antibodies were performed either as described by their manufacturer

(Abbott, N. Chicago, IL, Genelabs Diagnostics, Singapore) or using standard techniques known in the art (Harlow, et al.).

5

EXAMPLE 1CONSTRUCTION OF PNF2161 cDNA LIBRARIES

## A. ISOLATION OF RNA FROM SERA.

One milliliter of undiluted PNF 2161 serum was precipitated by the addition of PEG (MW 6,000) to 8% and 10 centrifugation at 12K, for 15 minutes in a microfuge, at 4°C. RNA was extracted from the resulting serum pellet essentially as described by Chomczynski.

The pellet was treated with a solution containing 4M guanidinium isothiocyanate, 0.18% 2- mercaptoethanol, and 15 0.5% sarcosyl. The treated pellet was extracted several times with acidic phenol-chloroform, and the RNA was precipitated with ethanol. This solution was held at -70°C for approximately 10 minutes and then spun in a microfuge at 4°C for 10 minutes. The resulting pellet was 20 resuspended in 100 µl of DEPC-treated (diethyl pyro-carbonate) water, and 10 µl of 3M NaOAc, pH = 5.2, two volumes of 100% ethanol and one volume of 100% isopropanol were added to the solution. The solution was held at -70°C for at least 10 minutes. The RNA pellet was recovered by centrifugation in a microfuge at 12,000 × g for 15 25 minutes at 5°C. The pellet was washed in 70% ethanol and dried under vacuum.

## B. SYNTHESIS OF cDNA

## 30 (i) FIRST STRAND SYNTHESIS

The synthesis of cDNA molecules was accomplished as follows. The above described RNA preparations were transcribed into cDNA, according to the method of Gubler et al. using random nucleotide hexamer primers (cDNA Synthesis Kit, BMB, Indianapolis, IN or GIBCO/BRL). 35

After the second-strand cDNA synthesis, T4 DNA polymerase was added to the mixture to maximize the number of

blunt-ends of cDNA molecules. The reaction mixture was incubated at room temperature for 10 minutes. The reaction mixture was extracted with phenol/chloroform and chloroform isoamyl alcohol.

5 The cDNA was precipitated by the addition of two volumes of 100% ethanol and chilling at -70°C for 15 minutes. The cDNA was collected by centrifugation, the pellet washed with 70% ethanol and dried under vacuum.

10 C. AMPLIFICATION OF THE DOUBLE STRANDED cDNA MOLECULES.

The cDNA pellet was resuspended in 12 µl distilled water. To the resuspended cDNA molecules the following components were added: 5 µl phosphorylated linkers (Linker AB, a double strand linker comprised of SEQ ID NO:1 and SEQ ID NO:2, where SEQ ID NO:2 is in a 3' to 5' orientation relative to SEQ ID NO:1 -- as a partially complementary sequence to SEQ ID NO:1), 2 µl 10× ligation buffer (0.66 M Tris.Cl pH=7.6, 50 mM MgCl<sub>2</sub>, 50 mM DTT, 10 mM ATP) and 1 µl T4 DNA ligase (0.3 to 0.6 Weiss Units).  
15 Typically, the cDNA and linker were mixed at a 1:100 ratio. The reaction was incubated at 14°C overnight. The following morning the reaction was incubated at 70°C for three minutes to inactivate the ligase.

To 100 µl of 10 mM Tris-Cl buffer, pH 8.3, containing 25 1.5 mM MgCl<sub>2</sub>, and 50 mM KCl (Buffer A) was added about 1 µl of the linker-ligated cDNA preparation, 2 µM of a primer having the sequence shown as SEQ ID NO:1, 200 µM each of dATP, dCTP, dGTP, and dTTP, and 2.5 units of *Thermus aquaticus* DNA polymerase (Taq polymerase). The reaction 30 mixture was heated to 94°C for 30 sec for denaturation, allowed to cool to 50°C for 30 sec for primer annealing, and then heated to 72°C for 0.5-3 minutes to allow for primer extension by Taq polymerase. The amplification reaction, involving successive heating, cooling, and 35 polymerase reaction, was repeated an additional 25-40 times with the aid of a Perkin-Elmer Cetus DNA thermal

cycler (Mullis; Mullis, et al.; Reyes, et al., 1991; Perkin-Elmer Cetus, Norwalk, CT).

After the amplification reactions, the solution was then phenol/chloroform, chloroform/isoamyl alcohol extracted and precipitated with two volumes of ethanol. The resulting amplified cDNA pellets were resuspended in 20  $\mu$ l TE (pH=7.5).

D. CLONING OF THE cDNA INTO LAMBDA VECTORS.

The linkers used in the construction of the cDNAs contained an EcoRI site which allowed for direct insertion of the amplified cDNAs into lambda gt11 vectors (Promega, Madison WI or Stratagene, La Jolla, CA). Lambda vectors were purchased from the manufacturer (Promega) which were already digested with EcoRI and treated with alkaline phosphatase, to remove the 5' phosphate and prevent self-ligation of the vector.

The EcoRI-digested cDNA preparations were ligated into lambda gt11 (Promega). The conditions of the ligation reactions were as follows: 1  $\mu$ l vector DNA (Promega, 0.5 mg/ml); 0.5 or 3  $\mu$ l of the PCR amplified insert cDNA; 0.5  $\mu$ l 10  $\times$  ligation buffer (0.5 M Tris-HCl, pH=7.8; 0.1 M MgCl<sub>2</sub>; 0.2 M DTT; 10 mM ATP; 0.5 g/ml bovine serum albumin (BSA)), 0.5  $\mu$ l T4 DNA ligase (0.3 to 0.6 Weiss units) and distilled water to a final reaction volume of 5  $\mu$ l.

The ligation reactions were incubated at 14°C overnight (12-18 hours). The ligated cDNA was packaged by standard procedures using a lambda DNA packaging system ("GIGAPAK", Stratagene, LaJolla, CA), and then plated at various dilutions to determine the titer. A standard X-gal blue/white assay was used to determine recombinant frequency of the libraries (Miller; Maniatis et al.).

Percent recombination in each library was also determined as follows. A number of random clones were selected and corresponding phage DNA isolated. Polymerase chain reaction (Mullis; Mullis, et al.) was then performed using isolated phage DNA as template and lambda DNA

sequences, derived from lambda sequences flanking the EcoRI ins rt site for the cDNA molecules, as primers. The presence or absence of insert was evident from gel analysis of the polymerase chain reaction products.

- 5       The cDNA-insert phage libraries generated from serum sample PNF 2161 was deposited with the American Type Culture Collection, 12301 Parklawn Dr., Rockville MD 20852, and has been assigned the deposit designation ATCC 75268 (PNF 2161 cDNA source).

10

EXAMPLE 2

IMMUNOSCREENING OF RECOMBINANT LIBRARIES

- The lambda gt11 libraries generated in Example 1 were immunoscreened for the production of antigens recognizable 15 by the PNF 2161 serum from which the libraries were generated. The phage were plated for plaque formation using the *Escherichia coli* bacterial plating strain *E. coli* KM392. Alternatively, *E. coli* Y1090R- may be used (Promega, Madison WI).

- 20       The fusion proteins expressed by the lambda gt11 clones were screened with serum antibodies essentially as described by Ausubel, et al.

Each library was plated at approximately  $2 \times 10^4$  phages per 150 mm plate. Plates were overlaid with 25 nitrocellulose filters overnight. Filters were washed with TBS (10 mM, Tris pH 7.5; 150 mM NaCl), blocked with AIB (TBS buffer with 1% gelatin) and incubated with a primary antibody diluted 100 times in AIB.

After washing with TBS, filters were incubated with a 30 second antibody, goat-anti-human IgG conjugated to alkaline phosphatase (Promega). Reactive plaques were developed with a substrate (for example, BCIP, 5-bromo-4-chloro-3-indolyl-phosphate), with NBT (nitro blue tetrazolium salt (Sigma)). Positive areas from the 35 primary screening were replated and immunoscreened until pure plaques were obtained.

EXAMPLE 3SCREENING OF THE PNF 2161 LIBRARY

The cDNA library of PNF 2161 in lambda gt11 was screened, as described in Example 2, with PNF 2161 sera.

5 The results of the screening are presented in Table 4.

Table 4  
PNF2161 Libraries

Library <sup>1</sup>	% Recomb. <sup>2</sup>	Antibody <sup>3</sup>	# Screened	# Clones Plaque-Purified
PNF/RNA	85	PNF	$5.5 \times 10^5$	4
PNF/RNA	90	PNF	$8 \times 10^4$	7
TOTALS:				11

- 10 15 1. cDNA library constructed from the indicated human source.
- 20 2. Percent recombinant clones in the indicated λgt11 library as determined by blue/white plaque assay and confirmed by PCR amplification of randomly selected clones.
- 25 3. Antisera source used for the immunoscreening of each indicated library.

One of the clones isolated by the above screen (PNF 2161 clone 470-20-1, SEQ ID NO:3; β-galactosidase in-frame fusion translated sequence, SEQ ID NO:4), was used to generate extension clones, as described in Example 6. The 30 clone 470-20-1 is deposited at Genelabs Technologies, Incorporated, 505 Penobscot Drive, Redwood City, CA 94063. Clone 470-20-1 nucleic acid sequence is presented as SEQ ID NO:3 (protein sequence SEQ ID NO:4). The isolated nucleic acid sequence without the SISPA cloning 35 linkers is presented as SEQ ID NO:19 (protein SEQ ID NO:20).

EXAMPLE 4CHARACTERIZATION OF THE IMMUNOREACTIVE 470-20-1 CLONE

## A. SOUTHERN BLOT ANALYSIS OF IMMUNOREACTIVE CLONES.

- The inserts of immunoreactive clones were screened  
5 for their ability to hybridize to the following control  
DNA sources: normal human peripheral blood lymphocyte  
(purchased from Stanford University Blood Bank, Stanford,  
California) DNA, and *Escherichia coli* KM392 genomic DNA  
(Ausubel, et al.; Maniatis, et al.; Sambrook, et al.).  
10 Ten micrograms of human lymphocyte DNA and 2 micrograms of  
*E. coli* genomic DNA were digested with EcoRI and HindIII.  
The restriction digestion products were  
electrophoretically fractionated on an agarose gel  
(Ausubel, et al.) and transferred to nylon or  
15 nitrocellulose membranes (Schleicher and Schuell, Keene,  
NH) as per the manufacturer's instructions.

Probes from the immunoreactive clones were prepared  
as follows. Each clone was amplified using primers  
corresponding to lambda gt11 sequences that flank the  
20 EcoRI cloning site of the gt11 vector. Amplification was  
carried out by polymerase chain reactions utilizing each  
immunoreactive clone as template. The resulting  
amplification products were digested with EcoRI, the  
amplified fragments gel purified and eluted from the gel  
25 (Ausubel, et al.). The resulting amplified fragments,  
derived from the immunoreactive clones, were then random  
prime labelled using a commercially available kit (BMB)  
employing <sup>32</sup>P-dNTPs.

The random primed probes were then hybridized to the  
30 above-prepared nylon membrane to test for hybridization of  
the insert sequences to the control DNAs. The 470-20-1  
insert did not hybridize with any of the control DNAs.

As positive hybridization controls, a probe  
derivative from a human C-kappa gene fragment (Hieter) was  
35 used as single gene copy control for human DNA and a *E.*  
*coli* polymerase gene fragment was similarly used for *E.*  
*coli* DNA.

B. GENOMIC PCR.

PCR detection was developed first to verify exogenicity with respect to several genomic DNAs which could have been inadvertently cloned during library construction, then to test for the presence of the cloned sequence in the cloning source and related specimen materials. Several different types of specimens, including SISPA-amplified nucleic acids and nucleic acids extracted from the primary source, and nucleic acids extracted from related source materials (e.g., from animal passage studies), were tested.

The term "genomic PCR" refers to testing for the presence of specific sequences in genomic DNA from relevant organisms. For example, a genomic PCR for a 15 Mystax-derived clone would include genomic DNAs as follows:

1. human DNA (1 µg/rxn.)
2. Mystax DNA (0.1-1 µg/rxn.)
3. *E. coli* (10-100 ng/rxn.)
- 20 4. yeast (10-100 ng/rxn.)

Human and Mystax DNAs are tested, as the immediate and ultimate source for the agent. *E. coli* genomic DNA, as a frequent contaminant of commercial enzyme preparations, is tested. Yeast is also tested, as a 25 ubiquitous organism, whose DNA can contaminate reagents and thus, be cloned.

In addition, a negative control (i.e., buffer or water only), and positive controls to include approximately  $10^5$ c/rxn., are also amplified.

30 Amplification conditions vary, as may be determined for individual sequences, but follow closely the following standard PCR protocol: PCR was performed in reactions containing 10 mM Tris, pH 8.3, 50 mM KCl, 1.75 mM MgCl<sub>2</sub>, 1.0 uM each primer, 200 uM each dATP, dCTP, and dGTP, and 35 300 µM dUTP, 2.5 units Taq DNA polymerase, and 0.2 units uracil-N-glycosylase per 100 ul reaction. Cycling was for at least 1 minute at 94°C, followed by 30 to 40

repetitions of denaturation (92-94°C for 15 seconds), annealing (55-56°C for 30 seconds), and extension (72°C for 30 seconds). PCR reagents were assembled, and amplification reactions were constituted, in a specially-  
5 designated laboratory maintained free of amplified DNA.

As a further barrier to contamination by amplified sequences and thus compromise of the test by "false positives," the PCR was performed with dUTP replacing TTP, in order to render the amplified sequences biochemically  
10 distinguishable from native DNA. To enzymatically render unamplifiable any contaminating PCR product, the enzyme uracil-N-glycosylase was included in all genomic PCR reactions. Upon conclusion of thermal cycling, the reactions were held at 72°C to prevent renaturation of  
15 uracil-N-glycosylase and possible degradation of amplified U-containing sequences.

A "HOT START PCR" was performed, using standard techniques ("AMPLIWAX", Perkin-Elmer Biotechnology; alternatively, manual techniques were used), in order to  
20 make the above general protocol more robust for amplification of diverse sequences, which ideally require different amplification conditions for maximal sensitivity and specificity.

Detection of amplified DNA was performed by  
25 hybridization to specific oligonucleotide probes located internal to the two PCR primer sequences and having no or minimal overlap with the primers. In some cases, direct visualization of electrophoresed PCR products was performed, using ethidium bromide fluorescence, but probe  
30 hybridization was in each case also performed, to help ensure discrimination between specific and non-specific amplification products. Hybridization to radiolabelled probes in solution was followed by electrophoresis in 8-15% polyacrylamide gels (as appropriate to the size of the  
35 amplified sequence) and autoradiography.

Clone 470-20-1 was tested by genomic PCR, against human, *E. coli*, and yeast DNAs. No specific sequence was

detected in negative control reactions, nor in any genomic DNA which was tested, and  $10^5$  copies of DNA/reaction resulted in a readily-detectable signal. This sensitivity (i.e.,  $10^5$ /reaction) is adequate for detection of single-copy human sequences in reactions containing 1 ug total DNA, representing the DNA from approximately  $1.5 \times 10^5$  cells.

#### C. DIRECT SERUM PCR

Serum or other cloning source or related source materials were directly tested by PCR using primers from selected cloned sequences. In these experiments, HGV viral particles were directly precipitated from sera with polyethylene glycol (PEG), or, in the case of PNF and certain other sera, were pelleted by ultracentrifugation. For purification of RNA, the pelleted materials were dissolved in guanidinium thiocyanate and extracted by the acid guanidinium phenol technique (Chomczynski, et al.).

Alternatively, a modification of this method afforded through and implemented by the use of commercially available reagents, e.g., "TRIREAGENT" (Molecular Research Center, Cincinnati, OH) or "TRIZOL" (Life Technologies, Gaithersburg, MD), and associated protocols was used to isolate RNA. In addition, RNA suitable for PCR analysis was isolated directly from serum or other fluids containing virus, without prior concentration or pelleting of virus particles, through the use of "PURESCRIPT" reagents and protocols (Gentra Systems, Minneapolis, MN).

Isolated DNA was used directly as a template for the PCR. RNA was reverse transcribed using reverse transcriptase (Gibco/BRL), and the cDNA product was then used as a template for subsequent PCR amplification.

In the case of 470-20-1, nucleic acid from the equivalent of 20-50 ul of PNF serum was used as the input template into each RT-PCR or PCR reaction. Primers were designed based on the 470-20-1 sequence, as follows: 470-20-1-77F (SEQ ID NO:9) and 470-20-1-211R (SEQ ID NO:10).

Reverse transcription was performed using MMLV-RT (Gibco/BRL) and random hexamers (Promega) by incubation at room temperature for approximately 10 minutes, 42°C for 15 minutes, and 99°C for 5 minutes, with rapid cooling to 5 4°C. The synthesized cDNA was amplified directly, without purification, by PCR, in reactions containing 1.75 mM MgCl<sub>2</sub>, 0.2-1 μM each primer, 200 uM each dATP, dCTP, dGTP, and dTTP, and 2.5-5.0 units Taq DNA polymerase ("AMPLITAQ", Perkin-Elmer) per 100 ul reaction. Cycling 10 was for at least one minute at 94°C, followed by 40-45 repetitions of denaturation (94°C for 15 seconds for 10 cycles; 92°C or 94°C for 15 seconds for the succeeding cycles), annealing (55°C for 30 seconds), and extension (72°C for 30 seconds), in the "GENEAMP SYSTEM 9600" 15 thermal cycler (Perkin-Elmer) or comparable cycling conditions in other thermal cyclers (Perkin-Elmer; MJ Research, Watertown, MA).

Positive controls consisted of (i) previously amplified PCR product whose concentration was estimated 20 using the Hoechst 33258 fluorescence assay, (ii) purified plasmid DNA containing the DNA sequence of interest, or (iii) purified RNA transcripts derived from plasmid clones in which the DNA sequence of interest is disposed under the transcriptional control of phage RNA promoters 25 such as T7, T3, or SP6 and RNA prepared through the use of commercially available *in vitro* transcription kits. In addition, an aliquot of positive control DNA corresponding to approximately 10-100 copies/rxn. can be spiked into reactions containing nucleic acids extracted from the 30 cloning source specimen, as a control for the presence of inhibitors of DNA amplification reactions. Each separate extract was tested with at least one positive control.

Specific products were detected by hybridization to a specific oligonucleotide probe 470-20-1-152F (SEQ ID 35 NO:16), for confirmation of specificity. Hybridization of 10 ul of PCR product was performed in solution in 20 ul reactions containing approximately  $1 \times 10^6$  cpm of <sup>32</sup>P-

labelled 470-20-1-152F. Specific hybrids were detected following electrophoretic separation from unhybridized oligo in polyacrylamide gels, and autoradiography.

In addition to PNF, extracted nucleic acids from 5 normal serum was also reverse transcribed and amplified, using the "serum PCR" protocol sequence. No signal was detected in normal human serum. The specific signal in PNF serum was reproducibly detected in multiple extracts, with the 470-20-1 specific primers.

10

#### D. AMPLIFICATION FROM SISPA UNCLONED NUCLEIC ACIDS

SISPA (Sequence-Independent Single Primer Amplification) amplified cDNA was used as templates (Example 1). Sequence-specific primers designed from 15 selected cloned sequences were used to amplify DNA fragments of interest from the templates. Typically, the templates were the SISPA-amplified samples used in the cloning manipulations. For example, amplification primers 470-20-1-77F (SEQ ID NO:9) and 470-20-1-211R (SEQ ID NO:10) were selected from the clone 470-20-1 sequence (SEQ ID NO:3). These primers were used in amplification reactions with the SISPA-amplified PNF2161 cDNA as a template.

The identity of the amplified DNA fragments were 25 confirmed by (i) hybridization with the specific oligonucleotide probe 470-20-1-152F (SEQ ID NO:16), designed based on the 470-20-1 sequence (SEQ ID NO:3) and/or (ii) size. The probe used for DNA blot detection was labelled with digoxigenin using terminal transferase 30 according to the manufacturer's recommendations (BMB). Hybridization to the amplified DNA was then performed using either Southern blot or liquid hybridization (Kumar, et al., 1989) analyses.

Positive control DNA used in the amplification 35 reactions was previously amplified PCR product whose concentration was estimated by the Hoechst 33258

fluorescence assay, or, alternatively, purified plasmid DNA containing the cloned inserts of interest.

The 470-20-1 specific signal was detected in cDNA amplified by PCR from SISPA-amplified PNF2161. Negative 5 control reactions were nonreactive, and positive control DNA templates were detected.

#### E. AMPLIFICATION FROM LIVER RNA SAMPLES.

RNA was prepared from liver biopsy material following 10 the methods of Cathal, et al., wherein tissue was extracted in 5M guanidine thiocyanate followed by direct precipitation of RNA by 4M LiCl. After washing of the RNA pellet with 2M LiCl, residual contaminating protein was removed by extraction with phenol:chloroform and the RNA 15 recovered by ethanol precipitation.

The 470-20-1 specific primers were also used in amplification reactions with the following RNA sources as substrate: normal mystax liver RNA, normal tamarin (*Sanguinus labiatus*) liver RNA, and MY131 liver RNA. 20 MY131 is a mystax that was infected with PNF 2161 plasma. Mystax 131 liver RNA did not give amplified products with the non-coding primers (SEQ ID NO:7 and SEQ ID NO:8) of HCV.

The amplification reactions were carried out in 25 duplicate for two experiments. The results of these amplification reactions are presented in Table 5.

Table 5

30 PCR with 470-20-1 Primers

		Exp. 1		Exp. 2	
		A	B	A	B
	Normal My liver RNA	-	-	-	-
	Normal tamarin liver RNA	-	-	-	-
35	My131 liver RNA	+	+	+	+
	PNF 2161	++	++	++	++

These results demonstrate the 470-20-1 sequences are present in the parent serum sample (PNF 2161) and in a liver RNA sample from a passage animal of the PNF 2161 sample (MY131). However, both control RNAs were negative  
5 for the presence of 470-20-1 sequences.

F. SCREENING OF A SERUM PANEL FOR HGV SEQUENCES BY POLYMERASE CHAIN REACTION USING RNA TEMPLATES.

1. PCR SCREENING OF HIGH-ALT DONORS FOR HGV

10 The disease association between HGV and liver disease was assessed by polymerase chain reaction screening, using HGV specific primers, of sera from hepatitis patients and from blood donors with abnormal liver function. The latter consisted of serum from blood donations with serum  
15 ALT levels greater than 45 International Units per ml.

A serum panel consisting of 152 total sera was selected. The following sera were selected for the serum panel: 104 high-ALT sera from screened blood donations at the Stanford University Blood Bank (SUBB); 34 N-(ABCDE) hepatitis sera from northern California, Egypt, and Peru; and 14 sera from other N-(ABCDE) donors suspected of having liver disease and/or hepatitis virus infection.  
20 The negative controls for the panel were as follows: 9 highly-screened blood donors (SUBB) notable for the absence of risk factors for viral infections ("supernormal" sera, e.g., O-negative, Rh-negative; negative for HIV, known hepatitis agents, and CMV; whose multiple previous blood donations had been transfused without causing disease); and 2 random blood donors.  
25 These sera were assayed for the presence of HGV specific sequences by RT-PCR using the 470-20-1 primers 77F (SEQ ID NO:9) and 211R (SEQ ID NO:10).

RNA extraction and RT-PCR were performed essentially as described in Example 4C, except that the primer 470-20-  
35 1-211R was 5'-biotinylated to facilitate rapid screening of amplified products by a method involving hybridization in solution, followed by affinity capture of hybridized

probe using streptavidin-coated paramagnetic beads. Methods for the analysis of nucleic acids by hybridization to specific labelled probes with capture of the hybridized sequences through affinity interactions are well known in 5 the art of nucleic acid analysis.

Depending on the amount of serum available for testing, RNA from 30 to 50  $\mu$ l of serum was used per RT/PCR reaction. Each serum was tested in duplicate, with positive controls corresponding to 10, 100, or 1000 copies 10 of RNA transcript per reaction and with appropriate negative (buffer) controls. No negative controls were reactive, and at least 10 copies per reaction were detectable in each PCR run. Indeterminate results were defined as specific hybridizing signal being present in 15 only one of two duplicate reactions.

Efficient, highly sensitive analysis of the products from the amplification analysis of this serum panel was performed using an instrument specifically designed for affinity-based hybrid capture using electrochemiluminescent 20 oligonucleotide probes (QPCR System 5000™, Perkin-Elmer). Assays utilizing the QPCR 5000™ have been described (DiCesare, et al; Wages, et al).

The products of each reaction were assayed by hybridization to probe 470-20-1-152F (5'-end-labelled with 25 an electrochemiluminescent ruthenium chelate), and measurement using the "QPCR 5000." Based on a cutoff of the sum of the mean and three times the standard deviation of negative controls in a given amplification run, a total of 34 possible positives were selected for confirmatory 30 testing.

The 34 samples were analyzed by solution hybridization and electrophoresis (Example 4C). Out of these 34 samples, 6 sera (i.e., 6/152) were shown to have specific hybridizing sequences in duplicate reactions. Of 35 these six samples, three were strongly reactive by comparison with positive controls: one High-ALT serum from SUBB, and two N-(ABCDE) sera from Egypt.

A second blood sample was obtained from the highly positive SUBB serum donor one year after the initial sample was taken. The second serum sample was confirmed to be HGV positive by the PCR methods described above.

- 5 This result confirms persistant infection by HGV in a human. The serum was designated "JC." Further, the serum donor was HCV negative and antibody negative for HAV and HBV.

In addition, a third N-(ABCDE) serum from Egypt, a  
10 northern California blood donor with N-(ABCDE) hepatitis, and a N-(ABCDE) hepatitis serum, were also shown to be weakly positive by this method. Two other sera gave indeterminate results, defined as the presence of specific sequences in one of two amplification reactions.

15 Subsequent PCR analysis of replicate serum aliquots from these positive and indeterminate sera resulted in positive results in 6 of 8 sera tested and indeterminate results in the remaining 2 sera. Thus, the specific hybridizing signal was reproducibly detected in 8 of the  
20 152 serum samples tested.

In contrast, none of the random donor or highly-screened "supernormal" sera (total 11) was positive in either set of PCR analysis.

These results reinforce the disease association  
25 between HGV and liver disease.

Further testing of sera from High-ALT donors has yielded the following results. A total of 495 sera have been tested, in addition to the initial panel of 104 sera described above. Of these 495 specimens, 6 were  
30 identified as HGV positive using the primer pair 470-20-1-77F (SEQ ID NO:9) and 470-20-1-211R (SEQ ID NO:10). Positive scores are based on repeated reactivity in at least 2 separate reactions. Accordingly, a detection rate of approximately 1-2% has been observed (8 of 599 tested).

G. INFECTIVITY OF HGV IN PRIMATES.

Two chimpanzees (designated CH1323 and CH1356), six cynomolgus monkeys (CY143, CY8904, CY8908, CY8912, CY8917, and CH8918), and four Mystax (MY98, MY187, MY229, MY254) 5 subjects were inoculated with PNF 2161. Pre-inoculation and post-inoculation sera were monitored for ALT and for the presence of HGV RNA sequences (as determined by PCR screening, described above).

One cynomologous monkey (CY8904) showed a positive 10 RNA PCR result and one indeterminant result from a total of 17 separate blood draws. In one chimpanzee, designated CH1356, was sustained viremia observed by RNA PCR. As shown in Table 6, no significant ALT elevation was observed, and circulating virus was detected only at time 15 points considerably after inoculation. Viremia was observed at and following 118 days post-inoculation. Suggestive reactivity was also observed in the first post-inoculation time-point (8 days), which may indicate residual inoculum.

20

Table 6

ALT and PCR Results from CH1356 Following  
Inoculation with PNF 2161

25

30

35

Days Post-Inoculation	ALT <sup>*</sup>	HGV PCR
0	59	-
8	65	±
15	85	-
22	89	-
29	89	-
36	86	-
39	31	-
47	74	-
54	40	-
61	57	-
84	65	±

Days Post-Inoculation	ALT*	HGV PCR
5	89	+
	98	-
	118	+
	125	+
	134	+
	159	+
	610	ALT not available

average ALT base-line before  
inoculation was 100.

The data presented above indicate that HGV infection was established in experimental primate subjects.

15

#### H. CHARACTERIZATION OF THE VIRAL GENOME.

The isolation of 470-20-1 from a cDNA library (Example 1) suggests that the viral genome detected in PNF 2161 is RNA. Further experiments to confirm the identity 20 of the HGV viral genome as RNA include the following.

Selective degradation of either RNA or DNA (e.g., by DNase-free RNase or RNase-free DNase) in the original cloning source followed by amplification with HGV specific primers and detection of the amplification products serves 25 to distinguish RNA from DNA templates.

An alternative method makes use of amplification reactions (nucleic acids from the original cloning source as template and HGV specific primers) that employ (i) a DNA-dependent DNA polymerase, in the absence of any RNA-dependent DNA polymerase (i.e., reverse transcriptase) in the reactions, and (ii) a DNA-dependent DNA polymerase and an RNA-dependent DNA polymerase in the reactions. In this 30 method, if the HGV genome is DNA or has a DNA intermediate, then amplified product is detected in both

types of amplification reactions. If the HGV genome is only RNA, the amplified product is detected in only the reverse transcriptase-containing reactions.

Total nucleic acid (i.e., DNA or RNA) was extracted from PNF 2161, using proteinase K and SDS followed by phenol extraction, as described in Example 4C. The purified nucleic acid was then amplified using polymerase chain reaction (PCR) where either (i) the PCR was preceded by a reverse transcription step, or (ii) the reverse transcription step was omitted. Amplification was reproducibly obtained only when the PCR reactions were preceded by reverse transcription. As a control, DNA templates were successfully amplified in separate reactions. These results demonstrate that the nature of the HGV viral genome is RNA.

The strand of the cloned, double-stranded DNA sequence that was originally present in PNF 2161 may be deduced by various means, including the following. Northern or dot blotting of the unamplified genomic RNA from an infected source serum can be performed, followed by hybridization of duplicate blots to probes corresponding to each strand of the cloned sequence. Alternatively, single-stranded cDNA probes isolated from M13 vectors (Messing), or multiple strand-specific oligonucleotide probes are used for added sensitivity. If the source serum contains single-stranded RNA, only one probe (i.e., sequences from one strand of the 470-20-1 clones) yield a signal, under appropriate conditions of hybridization stringency. If the source serum contains double-stranded RNA, both strand-probes will yield a signal.

The polymerase chain reaction, prefaced by reverse transcription using one or the other specific primer, represents a much more sensitive alternative to Northern blotting. Genomic RNA extracted from purified virions present in PNF 2161 serum is used as the input template into each RT/PCR. Rather than cDNA synthesis with random

hexamers, HGV sequence-specific primers were used. One cDNA synthesis reaction was performed with a primer complementary to one strand of the cloned sequence (e.g., 470-20-1-77F); a second cDNA synthesis reaction was also 5 performed using a primer derived from the opposite strand (e.g., 470-20-1-211R).

The resulting first strand cDNA was amplified in using two HGV specific primers. Controls were included for successful amplification by PCR (e.g., DNA controls). 10 RNA transcripts from each strand of the cloned sequence was also used, to control also for the reverse transcription efficiency obtained when using the specific primers which are described.

Specific products were detected by agarose gel 15 electrophoresis with ethidium bromide staining. DNA controls (i.e., double-stranded DNA controls for the PCR amplification) were successfully amplified regardless of the primer used for reverse transcription. Single-stranded RNA transcripts (i.e., controls for reverse 20 transcription efficiency and strand specificity) were amplified only when the opposite-strand primer was used for cDNA synthesis.

The PNF-derived HGV polynucleotide gave rise to a specific amplified product only when the primer 470-20-1-25 211R was used for reverse transcription, thus indicating that the original HGV polynucleotide sequence present in the serum is complementary to 470-20-1-211R and is likely a single-strand RNA.

30

EXAMPLE 5SUCROSE DENSITY GRADIENT SEPARATION OF PNF2161

## A. BANDING OF PNF-2161 AGENT.

A continuous gradient of 10-60% sucrose ("ULTRAPURE", Gibco/BRL) in TNE (50 mM Tris-Cl, pH 7.5, 100 mM NaCl, 1 35 mM EDTA) was prepared using a gradient maker from Hoefer Scientific (San Francisco, CA). Approximately 12.5 ml of the gradient was overlaid with 0.4 ml of PNF serum which

had been stored at -70°C, rapidly thawed at 37°C, then diluted in TNE.

The gradient was then centrifuged in the SW40 rotor (Beckman Instruments) at 40,000 rpm (approximately 200,000 5  $\times g$  at  $r_{av}$ ) at 4°C for approximately 18 hours. Fractions of volume approximately 0.6 ml were collected from the bottom of the tube, and 0.5 ml was weighed directly into the ultracentrifuge tube, for calculation of density.

10

Table 7

Measured Densities of PNF Fractions  
and Presence of 470-20-1

15

Fraction	Density	470-20-1 Detected*
1	1.274	-
2	1.274	-
3	1.266	-
4	1.266	-
5	1.260	-
6	1.254	-
7	1.248	+
8	1.206	+
9	1.146	+
10	1.126	+++
11	1.098	++++
12	1.068	+++
13	1.050	+
14	1.034	+
15	1.036	+
16	1.018	-
17	1.008	+
18	1.020	+

20

25

30

35

\* "+" and "-" scores were initially based on 40-cycle PCR. In order to distinguish "+", "++", "+++", and "++++", fractions giving initial positive scores (7-18) were amplified with 30 cycles of PCR.

The putative viral particles were then pelleted by centrifugation at 40,000 rpm in the Ti70.1 rotor (approximately 110,000 × g) at 4°C for 2 hours, and RNA was extracted using the acid guanidinium phenol technique 5 ("TRI REAGENT", Molecular Research Center, Cincinnati, OH), and alcohol-precipitated using glycogen as a carrier to improve recovery. The purified nucleic acid was dissolved in an RNase-free buffer containing 2 mM DTT and 1 U/μl recombinant RNasin.

10 Analysis of the gradient fractions by RNA PCR (Example 4C) showed a distinct peak in the 470-20-1 specific signal, localized in fractions of density ranging from 1.126 to 1.068 g/ml (Table 7). The 470-20-1 signal was thus shown, under these conditions, to form a discrete 15 band, consistent with the expected behavior of a viral particle in a sucrose gradient.

#### B. RELATIVE VIRAL PARTICLE DENSITIES.

PNF 2161 has been demonstrated to be co-infected with 20 HCV (see above). In order to compare the properties of the 470-20-1 viral particle to other known hepatitis viral particles, the serum PNF 2161 and a sample of purified Hepatitis A Virus were layered on a sucrose gradient (as described above). Fractions (0.6 ml) were collected, 25 pelleted and the RNA extracted. The isolated RNA from each fraction was subjected to amplification reactions (PCR) using HAV (SEQ ID NO:5; SEQ ID NO:6), HCV (SEQ ID NO:7; SEQ ID NO:8) and 470-20-1 (SEQ ID NO:9, SEQ ID NO:10) specific primers.

30 Product bands were identified by electrophoretic separation of the amplification reactions on agarose gels followed by ethidium bromide staining. The results of this analysis are presented in Table 8.

Table 8

	Average Density	HAV	HCV	470-20-1
5	1.269	-	-	-
	1.263	+	-	-
	1.260	+	-	-
	1.246	++	-	-
	1.238	++	-	-
	1.240	+	-	-
10	1.207	+	-	-
	1.193	+	-	-
	1.172	+	±	-
	1.150	+	±	±
	1.134	+	+	±
	1.118	+	+	+
15	1.103	+	+	+
	1.118	+	+	+
	1.103	+	+	+
	1.088	±	+	+
	1.084	-	+	+
	1.080	-	+	+
20	1.070	-	+	+
	1.057	-	+	±
	1.035	-	±	-
	1.017	-	-	-
	1.009	-	-	-

These results suggest that 470-20-1 particles are  
30 more similar to HCV particles than to HAV.

Further, serum PNF 2161 and HAV particles were  
treated with chloroform before sucrose gradient  
centrifugation. The results of these experiments suggest  
that 470-20-1 agent may be an enveloped virus since it has

more similar properties to an enveloped Flaviviridae member (HCV) than a non-enveloped virus (HAV).

EXAMPLE 6

5       GENERATION OF 470-20-1 EXTENSION CLONES

RNA was extracted directly from PNF2161 serum as described in Example 1. The RNA was passed through a "CHROMA SPIN" 100 gel filtration column (Clontech) to remove small molecular weight impurities. cDNA was 10 synthesized using a BMB cDNA synthesis kit. After cDNA synthesis, the PNF cDNA was ligated to a 50 to 100 fold excess of KL-1/KL-2 SISPA or JML-A/JML-B linkers (SEQ ID NO:11/SEQ ID NO:12, and SEQ ID NO:17/SEQ ID NO:18, respectively) and amplified for 35 cycles using either the 15 primer KL-1 or the primer JML-A.

The 470 extension clones were generated by anchored PCR of a 1  $\mu$ l aliquot from a 10  $\mu$ l ligation reaction containing EcoRI digested (dephosphorylated) lambda gt11 arms (1  $\mu$ g) and EcoRI digested PNF cDNA (0.2  $\mu$ g). PCR 20 amplification (40 cycles) of the ligation reaction was carried out using the lambda gt11 reverse primer (SEQ ID NO:13) in combination with either 470-20-77F (SEQ ID NO:9) or 470-20-1-211R (SEQ ID NO:10). All primer concentrations for PCR were 0.2  $\mu$ M.

25       The amplification products (9  $\mu$ l/100  $\mu$ l) were separated on a 1.5% agarose gel, blotted to "NYTRAN" (Schleicher and Schuell, Keene, NH), and probed with a digoxigenin labelled oligonucleotide probe specific for 470-20-1. The digoxigenin labeling was performed 30 according to the manufacturer's recommendations using terminal transferase (BMB). Bands that hybridized were gel-purified, cloned into the "TA CLONING VECTOR pCR II" (Invitrogen), and sequenced.

Sequencing was carried out using "DYEDEOXY TERMINATOR 35 CYCLE SEQUENCING" (a modification of the procedure of Sanger, et al.) on an Applied Biosystems model 373A DNA sequencing system according to the manufacturer's

- recommendations (Applied Biosystems, Foster City, CA). Sequence data is presented in the Sequence Listing. Sequences were compared with "GENBANK", EMBL database and dbEST (National Library of Medicine) sequences at both 5 nucleic acid and amino acid levels. Search programs FASTA, BLASTP, BLASTN and BLASTX (Altschul, et al.) indicated that these sequences were novel as both nucleic acid and amino acid sequences.
- Numerous clones having both 5' and 3' extensions to 10 470-20-1 were identified. All sequences are based on a consensus sequence from the sequencing of at least two independent isolates. This Anchor PCR approach was repeated in a similar manner to obtain further 5' and 3' extension sequences. These PCR amplification reactions 15 were carried out using the lambda gt11 reverse primer (SEQ ID NO:13) in combination with HGV specific primers derived from sequences obtained from previous extension clones. The substrate for these reactions was unpackaged PNF 2161 2-cDNA source DNA.
- 20 The individual consensus sequences were aligned, overlapping sequences identified and 9391 base pairs of the HGV sequence are presented as SEQ ID NO:14. This sequence represents a continuous open reading frame (SEQ ID NO:15).
- 25 The relationship between the original 470-20-1 clone and the sequences obtained by extension is shown schematically in Figure 1. As seen in the figure, the DNA strand having opposite polarity to the protein coding sequence of 470-20-1 comprising a long continuous open 30 reading frame.
- The amino acid sequence of HGV was compared against the sequences of all viral sequence in the PIR database (IntelliGenetics, Inc., Mountain View, CA) of protein sequences. The comparison was carried out using the 35 "SSEARCH" program of the "FASTA" suite of programs version 1.7 (Pearson, et al.). Regions of local sequence similarities were found between the HGV sequences and two

virus's in the Flaviviridae family of viruses. The similarity alignments are presented in Figures 5A and 5B.

Present in these alignments are motifs for the RNA dependent RNA polymerase (RDRP) of these viruses.

- 5 Conserved RDRP amino acid motifs are indicated in Figures 5A and 5B by stars and uppercase, bold letters (Koonin and Dolja). These alignments demonstrate that this portion of the HGV coding sequence correspond to RDRP. This alignment data combined with the data concerning the RNA  
10 genome of HGV supports the placement of HGV as a member of the Flaviviridae family.

- The global amino acid sequence identities of the HGV polyprotein (SEQ ID NO:15) with HoCV (Hog Cholera Virus) and HCV are 17.1% and 25.5%, respectively. Such levels of  
15 global sequence identity demonstrates that HGV is a separate viral entity from both HoCV and HCV. To illustrate, in two members of the Flaviviridae family of viruses BVDV (Bovine Diarrhea Virus) and HCV, 16.2% of the amino acids can be globally aligned with HGV.

- 20 Members within a genus generally show high homology when aligned globally, for example, BVDV vs. HoCV show 71.2% identity. Various members (variants) of the unnamed genus of which HCV is a member are between 65% and 100% identical when globally aligned.

25

EXAMPLE 7

ISOLATION OF 470-20-1 FUSION PROTEIN

A. EXPRESSION AND PURIFICATION OF 470-20-1/GLUTATHIONE-S-TRANSFERASE FUSION PROTEIN

- 30 Expression of a glutathione-S-transferase (sj26) fused protein containing the 470-20-1 peptide was achieved as follows. A 237 base pair insert (containing 17 nucleotides of SISPA linkers on both sides) corresponding to the original lambda gt11 470-20-1 clone was isolated  
35 from the lambda gt11 470-20-1 clone by polymerase chain reaction using primers gt11 F(SEQ ID NO:25) and gt11 R(SEQ ID NO:13) followed by Eco RI digestion.

The insert was cloned into a modified pGEX vector, pGEX MOV. pGEX MOV encodes sj26 protein fused with six histidines at the carboxy terminal end (sj26his). The 470-20-1 polypeptide coding sequences were introduced into 5 the vector at a cloning site located downstream of sj26his coding sequence in the vector. Thus, the 470-20-1 polypeptide is expressed as sj26his/470-20-1 fusion protein. The sj26 protein and six histidine region of the 10 fusion protein allow the affinity purification of the fusion protein by dual chromatographic methods employing glutathione-conjugated beads (Smith, D.B., et al.) and immobilized metal ion beads (Hochula; Porath).

E. coli strain W3110 (ATCC catalogue number 27352) was transformed with pGEX MOV and pGEX MOV containing 470-15 20-1 insert. Sj26his protein and 470-20-1 fusion protein were induced by the addition of 2 mM isopropyl- $\beta$ -thiogalactopyranoside (IPTG). The fusion proteins were purified either by glutathione-affinity chromatography or by immobilized metal ion chromatography (IMAC) according 20 to the published methods (Smith, D.B., et al.; Porath) in conjunction with conventional ion-exchange chromatography.

The purified 470-20-1 fusion protein was immunoreactive with PNF 2161. However, purified sj26his protein was not immunoreactive with PNF 2161, indicating 25 the presence of specific immunoreaction between the 470-20-1 peptide and PNF 2161.

B. ISOLATION OF 470-20-1/B-GALACTOSIDASE FUSION PROTEIN  
KM392 lysogens infected either with lambda phage gt11 30 or with gt11/470-20-1 are incubated in 32°C until the culture reaches to an O.D. of 0.4. Then the culture is incubated in a 43°C water bath for 15 minutes to induce gt11 peptide synthesis, and further incubated at 37°C for 1 hour. Bacterial cells are pelleted and lysed in lysis 35 buffer (10 mM Tris, pH 7.4, 2 % "TRITON X-100" and 1% aprotinin). Bacterial lysates are clarified by centrifugation (10K, for 10 minutes, Sorvall JA20 rotor)

and the clarified lysates are incubated with Sepharose 4B beads conjugated with anti- $\beta$ -galactosidase (Promega).

Binding and elution of  $\beta$ -galactosidase fusion proteins are performed according to the manufacturer's instruction. Typically binding of the proteins and washing of the column are done with lysis buffer. Bound proteins are eluted with 0.1 M carbonate/bicarbonate buffer, pH 10. The purified 470-20-1/b-galactosidase protein is immunoreactive with both PNF2161 and anti- $\beta$ -galactosidase antibody. However,  $\beta$ -galactosidase, expressed by gt11 lysogen and purified, is not immunoreactive with PNF2161 but immunoreactive with anti- $\beta$ -galactosidase antibody.

15

#### EXAMPLE 8

##### PURIFICATION OF THE 470-20-1 FUSION PROTEIN AND PREPARATION OF ANTI-470-20-1 ANTIBODY

###### A. GLUTATHIONE AFFINITY PURIFICATION

Materials included 50 ml glutathione affinity matrix reduced form (Sigma), XK 26/30 Pharmacia column, 2.5 x 10 cm Bio-Rad "ECONO-COLUMN" (Richmond, CA), Gilson (Middleton, WI) HPLC, DTT (Sigma), glutathione reduced form (Sigma), urea, and sodium phosphate dibasic.

The following solutions were used in purification of 25 the fusion protein:

Buffer A: phosphate buffer saline, pH 7.4, and

Buffer B: 50 mM Tris Ph 8.5, 8 mM glutathione,

(reduced form glutathione)

Strip buffer: 8 M urea, 100 mM Tris pH 8.8, 10 mM 30 glutathione, 1.5 NaCl.

*E. coli* carrying the plasmid pGEX MOV containing 470-20-1 insert, were grown in a fermentor (20 liters). The bacteria were collected and lysed in phosphate buffered 35 saline (PBS) containing 2 mM phenylmethyl sulfonyl fluoride (PMSF) using a micro-fluidizer. Unless otherwise

noted, all of the following procedures were carried out at 4°C.

The crude lysate was prepared for loading by placing lysed bacteria into "OAKRIDGE" tubes and spinning at 20K 5 rpms (40k x g) in a Beckman model JA-20 rotor. The supernatant was filtered through a 0.4 µm filter and then through a 0.2 µm filter.

The 2.5 x 10 cm "ECONO-COLUMN" was packed with the glutathione affinity matrix that was swelled in PBS for 10 two hours at room temperature. The column was brought into equilibrium by washing with 4 bed volumes of PBS.

The column was loaded with the crude lysate at a flow rate of 8 ml per minute. Subsequently, the column was washed with 5 column volumes of PBS at the same flow rate.

15 The column was eluted by setting the flow rate to 0.75-1 ml/min. and introducing Buffer B. Buffer B was pumped through the column for 5 column volumes and two-minute fractions were collected. An exemplary elution profile is shown in Figure 2. The content and purity of 20 the proteins present in the fractions were assessed by standard SDS PAGE (Figure 3). The 470-20-1/sj26his fusion protein was identified based on its predicted molecular weight and its immunoreactivity to PNF 2161 serum. For further manipulations, the protein can be isolated from 25 fractions containing the fusion protein or from the gel by extraction of gel regions containing the fusion protein.

B. PURIFICATION OF CLONE 470-20-1 FUSION PROTEIN BY ANION EXCHANGE.

30 Solutions include the following:

Buffer A (10 mM sodium phosphate pH 8.0, 4 M urea, 10 mM DTT);

Buffer B (10 mM sodium phosphate pH 8.0, 4 M urea, 10 mM DTT, 2.0 M NaCl); and

35 Strip Buffer (8 M urea, 100 mM Tris pH 8.8, 10 mM glutathione, 1.5 NaCl).

Crude lysate (or other protein source, such as pooled fractions from above) was loaded onto "HIGH-Q-50" (Biorad, Richmond, CA) column at a flow rate of 4.0 ml/min. The column was then washed with Buffer A for 5 column volumes

5. at a flow rate of 4.0 ml/min.

- After these washes, a gradient was started and ran from Buffer A to Buffer B in 15 column volumes. The gradient then stepped to 100% Buffer B for one column volume. An exemplary gradient is shown in Figure 4A.
- 10 Fractions were collected every 10 minutes. Purity of the 470-20-1/sj26his fusion protein was assessed by standard SDS-PAGE (Figures 4B and 4C) and relevant fractions were pooled (approximately fractions 34 through 37, Figure 4C).

15 C. PREPARATION OF ANTI-470-20-1 ANTIBODY

The purified 470-20-1/sj26his fusion protein is injected subcutaneously in Freund's adjuvant in a rabbit. Approximately 1 mg of fusion protein is injected at days 0 and 21, and rabbit serum is typically collected at 6 and 8 weeks.

A second rabbit is similarly immunized with purified sj26his protein.

Minilysates are prepared from bacteria expressing the 470-20-1/sj26his fusion protein, sj26his protein, and  $\beta$ -galactosidase/470-20-1 fusion protein. The lysates are fractionated on a gel and transferred to a membrane. Separate Western blots are performed using the sera from the two rabbits.

Serum from the animal immunized with 470-20-1 fusion protein is immunoreactive with all sj26his fusion protein in minilysates of IPTG induced E. coli W3110 that are transformed either with pGEX MOV or with pGEX MOV containing 470-20-1 insert. This serum is also immunoreactive with the fusion protein in the minilysate from the 470-20-1 lambda gt11 construct.

The second rabbit serum is immunoreactive with both sj26his and 470-20-1/sj26his fusion proteins in the

minilysates. This serum is not expected to immunoreactive with 470-20-1/ $\beta$ -galactosidase fusion protein in the minilysate from the 470-20-1 lambda gt11 construct. None of the sera are expected to be immunoreactive with  $\beta$ -

5 galactosidase.

Anti-470-20-1 antibody present in the sera from the animal immunized with the fusion protein is purified by affinity chromatography (using the 470-20-1 ligand).

Alternatively, the fusion protein can be cleaved to 10 provide the 470-20-1 antigen free of the sj-26 protein sequences. The 470-20-1 antigen alone is then used to generate antibodies as described above.

EXAMPLE 9

15

SEROLOGY

A. WESTERN BLOTT ANALYSIS OF SERA PANELS

The 470-20-1 fusion antigen (described above) was used to screen panels of sera. Many of the panels were of human sera derived both from individuals suffering from 20 hepatitis and uninfected controls.

Affinity purified 470-20-1 fusion antigen (Example 8) was loaded onto a 12% SDS-PAGE at 2  $\mu$ g/cm. The gel was run for two hours at 200V. The antigen was transferred from the gel to a nitrocellulose filter.

25 The membrane was then blocked for 2 hours using a solution of 1% bovine serum albumin, 3% normal goat serum, 0.25% gelatin, 100 mM NaPO<sub>4</sub>, 100 mM NaCl, and 1% nonfat dry milk. The membrane was then dried and cut into 1-2 mm strips; each strip contained the 470-20-1 fusion antigen. 30 The strip was typically rehydrated with TBS (150 mM NaCl; 20 mM Tris HCl, pH 7.5) and incubated in panel sera (1:100) overnight with rocking at room temperature.

The strips were washed twice for five minutes each time in TBS plus "TWEEN 20" (0.05%), and then washed twice 35 for five minutes each time in TBS. The strips were then incubated in secondary antibody (Promega anti-human IgG-Alkaline Phosphatase conjugate, 1:7500), for 1 hour with

rocking at room temperature. The strips were then washed twice × 5 minutes in TBS + "TWEEN 20", then twice × 5 minutes in TBS.

- Bound antibody was detected by incubating the strips
- 5 in a substrate solution containing BCIP (Example 2) and NBT (Example 2) in pH 9.5 buffer (100 mM Tris, 100 mM NaCl, 5 mM MgCl<sub>2</sub>). Color development was allowed to proceed for approximately 15 minutes at which point color development was halted by 3 washes in distilled H<sub>2</sub>O.
- 10 Test sera were derived from the following groups of individuals: (i) blood donors, negative for HBV Ab, surface Ag, negative for HCV, HIV, HTLV-1 Abs; (ii) HBV, sera from individuals who are infected with Hepatitis B virus; (iii) HCV, sera from individuals infected with
- 15 Hepatitis C virus by virtue of being reactive in a second-generation HCV ELISA assay; and (iv) HXV, individuals serologically negative for HAV, HBV, HCV, or HEV.

The results of these screens are presented in Table

9.

20

Table 9

470-20-1 Sera Panelling Result Summary

Sample	No. Human <sup>*</sup> Sera Tested	+	IND <sup>*</sup>	-
blood donor	30	1 (3.3%)	2 (6.7%)	27 (90.0%)
HBV	40	7 (17.5%)	4 (10.0%)	29 (72.5%)
HCV	38	11 (28.95%)	11 (28.95%)	16 (42.1%)
HXV	122	20 (16.4%)	12 (9.8%)	90 (73.8%)

\* Indeterminate, weak reactivity

- 35 These results suggest the presence of the 470-20-1 antigen in a number of different sera samples. The antigen is not immunoreactive with normal human sera.

B. GENERAL ELISA PROTOCOL FOR DETECTION OF ANTIBODIES

Polystyrene 96 well plates ("IMMULON II" (PGC)) are coated with 5 µg/ml (100 µL per well) antigen in 0.1 M sodium bicarbonate buffer, pH 9.5. Plates are sealed with "PARAFILM" and stored at 4°C overnight.

Plates are aspirated and blocked with 300 uL 10% normal goat serum and incubated at 37°C for 1 hr.

Plates are washed 5 times with PBS 0.5% "TWEEN-20".

Antisera is diluted in 1 x PBS, pH 7.2. The desired dilution(s) of antisera (0.1 mL) are added to each well and the plate incubated 1 hour at 37°C. The plates are then washed 5 times with PBS 0.5% "TWEEN-20".

Horseradish peroxidase (HRP) conjugated goat anti-human antiserum (Cappel) is diluted 1/5,000 in PBS. 0.1 mL of this solution is added to each well. The plate is incubated 30 min at 37°C, then washed 5 times with PBS.

Sigma ABTS (substrate) is prepared just prior to addition to the plate.

The reagent consists of 50 ml 0.05 M citric acid, pH 4.2, 0.078 ml 30% hydrogen peroxide solution and 15 mg ABTS. 0.1 mL of the substrate is added to each well, then incubated for 30 min at room temperature. The reaction is stopped with the addition of 0.050 mL 5% SDS (w/v). The relative absorbance is determined at 410 nm.

25

EXAMPLE 10

Preliminary Mapping of HGV Epitopes

An approximately 7.3 kb coding sequence of HGV was subcloned as 77 distinct but overlapping cDNA fragments. The length of most cDNA fragments ranged from about 200 bp to about 500 bp. The cDNA fragments were cloned separately into the expression vector, pGEX-HisB. This vector is similar to pGEX-MOV, described above.

pGEX-hisB is a modification of pGEX-2T (Genbank accession number A01438; a commercially available expression vector). The vector pGEX-2T has been modified by insertion of a NcoI site directly downstream from the

thrombin cleavage site. This site is followed by a *BamHI* site, which is followed by a poly-histidine (six histidines) encoding sequence, followed by the *EcoRI* site found in pGEX-2T. Coding sequences of interest are 5 typically inserted between the *NcoI* site and the *BamHI* site. In Figure 6 (SEQ ID NO:96), the inserted sequence encodes the GE3-2 antigen. The rest of the vector sequence is identical to pGEX-2T. Expression of fusion protein is carried out essentially as described above with 10 other pGEX-derived expression vectors.

Cloning of all 24 fragments was carried out essentially as described below, where specific primers were selected for each of the 24 coding regions. Typically, the 5' primer contained a *NcoI* restriction site 15 and the 3' primer contained a *BamHI* restriction site. The *NcoI* primers in the amplified fragments allowed in-frame fusion of amplified coding sequences to the GST-Sj26 coding sequence in the expression vector pGEX-Hisb. Expressed recombinant proteins were analyzed for specific 20 immunoreactivity against putative HGV-infected human sera by Western blot.

Two fragments designated GE3 and GE9 encoded antigens that gave a clear immunogenic response when reacted with putative HGV-infected human sera.

25

#### A. CLONING OF GE3, GE9, GE15, AND GE17.

The coding sequence inserts for clones GE3 and GE9 were generated by polymerase chain reaction from SISPA-amplified double-stranded cDNA or RNA obtained from PNF 30 2161, using PCR primers specific for each fragment.

In the GE3-5' primer (GE-3F, SEQ ID NO:30) a silent point mutation was introduced to modify a natural *NcoI* restriction site. The GE3-3' primer was GE-3R (SEQ ID NO:31). The GE9-5' primer was GE-9F (SEQ ID NO:32) and 35 the GE9-3' primer was GE-9R (SEQ ID NO:33). The GE15-5' primer was GE-15F (SEQ ID NO:92) and the GE15-3' primer was GE-15R (SEQ ID NO:93). The GE17-5' primer was GE-17F

(SEQ ID NO:94) and the GE17-3' primer was GE-17R (SEQ ID NO:95). Using these primers, PCR amplification products were generated. The amplification products were gel purified, digested with *NcoI* and *BamHI*, and gel purified again. The purified *NcoI/BamHI* GE3, GE9, GE15 and GE17 fragments were independently ligated into dephosphorylated, *NcoI/BamHI* cut pGEX-HisB vectors.

Each ligation mixture was transformed into *E.coli* W3110 strain and ampicillin resistant colonies were selected. The ampicillin resistant colonies were resuspended in a Tris/EDTA buffer were analyzed by PCR, using primers homologous to pGEX vector sequences flanking the inserted molecules, to confirm the presence of insert sequences. Four candidate clones were designated GE3-2 (SEQ ID NO:34), GE9-2 (SEQ ID NO:36), GE15-1 (SEQ ID NO:88) and GE17-2 (SEQ ID NO:90), respectively.

B. EXPRESSION OF THE GE3-2, GE9-2, GE15-1, AND GE17-2 FUSION PROTEINS.

Colonies of ampicillin resistant bacteria carrying GE3-2, GE9-2, GE15-1, and GE17-2 containing-vectors were individually inoculated into LB medium containing ampicillin. The cultures were grown to OD of 0.8 to 0.9 at which time IPTG (isopropylthio-beta-galactoside; Gibco-BRL) was added to a final concentration of 0.3 to 0.5 mM, for the induction of protein expression. Incubation in the presence of IPTG was continued for 3 to 4 hours.

Bacterial cells were harvested by centrifugation and resuspended in SDS sample buffer (0.0625 M Tris, pH 6.8, 30% glycerol, 5% mercaptoethanol, 2.3% SDS). The resuspended pellet was boiled for 5 min. and then cleared of cellular debris by centrifugation. The supernatants obtained from IPTG-induced cultures of GE3-2, GE9-2, GE15-1, and GE17-2 were analyzed by SDS-polyacrylamide gel electrophoresis (PAGE). The proteins from these gels were then transferred to nitrocellulose filters (i.e., by Western blotting). The filters were then exposed to PNF

2161, JC and supernormal serum. JC is the HGV-positive sera identified in Example 4F that was rejected by the blood bank for being High ALT. A second sample, taken one year after the initial serum sample, was also positive for 5 HGV by PCR analysis. Immunoreactivity of JC serum with bands at the appropriate molecular weight for the fusion proteins demonstrated the successful expression of the fusion protein by the bacterial cells.

The fusion proteins were purified from bacterial cell 10 lysates essentially as in Example 7 using dual chromatographic methods employing glutathione-conjugated beads (Smith, D.B., et al.) and immobilized metal ion beads (Hochula; Porath).

15 C. WESTERN BLOTTING ANALYSIS OF PURIFIED HGV PROTEINS.

Various amounts of the purified HGV proteins (e.g., GE3-2 and GE9-2 proteins) were loaded on 12% acrylamide gels. Following PAGE, proteins were transferred from the gels to nitrocellulose membranes, using standard procedures. Individual membranes were incubated with one of 20 a number of human or mouse sera. Excess sera were removed by washing the membranes.

These membranes were incubated with alkaline phosphatase-conjugated goat anti-human antibody (Promega) 25 or alkaline phosphatase-conjugated goat anti-mouse antibodies (Sigma), depending on the serum being used for screening. The membranes were washed again, to remove excess goat anti-human IgG antibody, and exposed to NBT/BCIP. Photographs of exemplary stained membranes 30 having the GE3 fusion protein are shown in Figures 7A to 7D.

The Figures show the results of Western blot analysis of the purified GE3-2 protein using the following sera: N-(ABCDE) human (JC) serum (Figure 7A), N-(ABDE) human 35 (PNF 2161) serum (Figure 7B), a sup r normal (SN2) serum (Figure 7C), and mouse monoclonal antibody (RM001) directed against GST-Sj26 protein (Figure 7D).

In each of the figures, lane 1 contains molecular weight standards, and lanes 2-5 contain, respectively, the following amounts of the GE3-2 fusion protein: 4 µg, 2 µg, 1 µg, and 0.5 µg. Numbers represent loading amounts 5 in micrograms per 0.6 centimeter of gel (well size). Dilutions of the human JC, PNF 2161 and Super Normal 2 sera were 1:100. The anti-sj26 dilution was 1:1000. The band seen at about 97K in the JC blot is reactivity against a minor contaminant in the GE3.2 fusion protein 10 preparation. Protein marker sizes are 142.9, 97.2, 50, 35.1, 29.7 and 21.9 KD.

As shown in Figures 7A to 7D, GE3-2 showed specific immunoreactivity with JC serum. GE3-2 reacted weakly with PNF 2161 serum and would be scored as an indeterminant or 15 negative.

In parallel experiments, GE9-2 showed weak but specific immunoreactivity toward PNF 2161 serum. Further, GE15-1 and GE-17 showed weak but specific immunoreactivity toward PNF 2161 and T55806. T55806 is human serum that 20 contains HGV; this sera was identified as HGV positive by PCR, as described in Example 4.

#### EXAMPLE 11

##### Construction of an Exemplary Epitope Library

25 Polymerase Chain Reactions were employed to amplify 3 overlapping DNA fragments from PNF 2161 SISPA-amplified cDNA. The PNF 2161 SISPA-amplified cDNA was prepared using the JML-A/B linkers (SEQ ID NO:38 and SEQ ID NO:39). One microliter of this material was re-amplified for 30 30 cycles (1 minute at 94°C, 1.5 minutes at 55°C and 2 minutes at 72°C) using 1 µM of the JML-A primers. The total reaction volume was 100 µl. The products from 3 of these amplifications were combined and separated from excess PCR primers by a single pass through a "WIZARD PCR 35 COLUMN" (Promega) following the manufacturer's instructions. The "WIZARD PCR COLUMN" is a silica based resin that binds DNA in high ionic strength buffers and

will release DNA in low ionic strength buffers. The amplified DNA was eluted from the column with 100  $\mu$ l distilled H<sub>2</sub>O.

The eluted DNA was fractionated on a 1.5% Agarose TBE gel (Maniatis, et al.) and visualized with UV light following ethidium bromide staining. A strong smear of DNA fragments between 150 and 1000 bp was observed. One microliter of the re-amplified cDNA was used as for template in PCR reactions with each primer pair presented in Table 10.

Table 10

Primers	SEQ ID NO:	Size of Amplified Fragment
470ep-F1 470ep-R1	SEQ ID NO:40 SEQ ID NO:41	810
470ep-F2 470ep-R3	SEQ ID NO:42 SEQ ID NO:43	750
470ep-F4 470ep-R4	SEQ ID NO:44 SEQ ID NO:45	669

15  
20

The primers were designed to result in the amplification of HGV specific DNA fragments of the sizes indicated in Table 10. In the amplification reactions, the primer pairs were used at a concentration of 1  $\mu$ M. Amplifications were for 30 cycles of 1 minute at 94, 1.5 minutes at 54°C and 3 minutes at 72°C in a total reaction volume of 100  $\mu$ l. Each of the three different primer pair PCR reactions resulted in the specific amplification of products having the expected sizes. For each primer pair reaction, amplification products from 3 independent PCR reactions were combined and purified using a "WIZARD PCR COLUMN" as described above. The purified products were eluted in 50  $\mu$ l dH<sub>2</sub>O.

35 Samples from each purified product (14  $\mu$ l, containing approximately 1 - 2  $\mu$ g of each primer-pair amplified DNA

fragment) were combined. The combined sample of all three different amplified fragments was added to 5  $\mu$ l of 10X DNase Digestion buffer (500 mM Tris PH 7.5, 100 mM MnCl<sub>2</sub>) and 2  $\mu$ l of dH2O. From this digestion mixture, a 10  $\mu$ l sample was removed and placed in a tube containing 5  $\mu$ l of Stop solution (100 mM EDTA, pH 8.0). This sample was the 0 "minutes of digestion" time point. The rest of the digestion reaction was placed at 25°C. To the digestion mixture 1  $\mu$ l of 1/25 diluted RNase-free DNase I 10 (Stratagene) was added. At various time points 10  $\mu$ l aliquots were withdrawn and mixed with 5  $\mu$ l of Stop solution. The DNaseI digested DNA products were analyzed on a 1.5% Agarose TBE gel.

The results of several digestion experiments showed 15 that 40 minutes of digestion provided a good distribution of DNA fragments in the size range of 100 - 300 bp. A DNase I digestion was then repeated with the entire digestion being left for 40 minutes at room temperature. The digestion was stopped by the addition of 18  $\mu$ l of Stop 20 Buffer and the digested DNA products were purified using a "WIZARD PCR COLUMN." The "WIZARD-PCR COLUMN" was eluted with 50  $\mu$ l of dH2O and the eluted DNA added to the following reaction mixture: 7  $\mu$ l of Restriction Enzyme Buffer C (Promega, 10 mM MgCl<sub>2</sub>, 1 mM DTT, 50 mM NaCl, 10 mM Tris, pH 7.9, 1X concentration); 11  $\mu$ l of 1.25 mM dNTPs; 25 and 2  $\mu$ l T4 DNA Polymerase (Boehringer-Mannheim). This reaction mixture was held at 37°C for 30 minutes, at which point 70  $\mu$ l of pH 8.0 phenol/CHCl<sub>3</sub> was added and mixed. The phenol/CHCl<sub>3</sub> was removed and extracted once to yield a 30 total aqueous volume of 150  $\mu$ l containing the DNA sample. The DNA was ethanol precipitated using 2 volumes of absolute ethanol and 0.5 volume of 7.5 M NH<sub>4</sub>-acetate. The DNA was pelleted by centrifugation for 15 minutes at 14,000 rpm in an "EPPENDORF MICROFUGE", dried for 5 35 minutes at 42°C and resuspended in 25  $\mu$ l of dH2O.

The DNA was ligated to 5' phosphorylated SISPA linkers KL1 (SEQ ID NO:46) and KL2 (SEQ ID NO:47).

Several different concentrations of SISPA linkers and DNA was tested. The highest level of ligation (assessed as described below) occurred under the following ligation reaction conditions: 6  $\mu$ l of DNA, 2  $\mu$ l of  $5.0 \times 10^{-12}$  M KL1/KL2 linkers, 1  $\mu$ l of 10X ligase buffer (New England Biolabs), and 1  $\mu$ l of 400 Units/ $\mu$ l T4 DNA Ligase (New England Biolabs) in a total reaction volume of 10  $\mu$ l. Ligations were carried out overnight at 16°C.

Two reactions were run in parallel as follows. A 2  $\mu$ l sample of the ligated material was amplified using the KL1 SISPA primer in a total reaction volume of 100  $\mu$ l (25 cycles of 1 minute at 94°C, 1.5 minutes at 55°C and 2 minutes at 72°C). The degree of ligation was assessed by separating 1/5 of the PCR reaction amplified products by electrophoresis using a 1.5% agarose TBE gel. The gel was stained with ethidium bromide and the bands visualized with UV light.

The amplification products from the duplicate reactions were purified using "WIZARD PCR COLUMNS" and the purified DNA eluted in 50  $\mu$ l of dH2O. A twenty-five microliter aliquot of the PCR KL1/KL2 amplified DNA was digested with 36 Units of EcoRI (Promega) in a total volume of 30  $\mu$ l. The reaction was carried out overnight at 37°C. The Digested DNA was purified using a "SEPHADEX G25" spin column.

The EcoRI digested DNA was ligated in overnight reactions to  $\lambda$ gt11 arms that were pre-digested with EcoRI and treated with calf intestinal alkaline phosphatase (Stratagene, La Jolla, CA). The ligation mixture was packaged using a "GIGAPACK GOLD PACKAGING EXTRACT" (Stratagene) following manufacturer's instructions.

Titration of the amount of recombinant phage obtained was performed by plating a 1/10 dilution of the packaged phage on a lawn of KM-392, where the plate contained 20  $\mu$ l of a 100 mg/ml solution of x-gal (5-Bromo-4-chloro-3-indolyl- $\beta$ -D-galactoside; Sigma) and 20  $\mu$ l of a 0.1 M solution of IPTG (Isopropyl-1-thio- $\beta$ -D-galactoside; Sigma). A titer

100

was obtained of  $1.2 \times 10^6$  phage/ml containing over 75% recombinant phage.

The percentage of recombinant plaques was confirmed by PCR analysis of 8 randomly picked plaques using primers 5 11F (SEQ ID NO:25) and 11R (SEQ ID NO:13). This packaged library containing the DNA fragments derived from the digestion of the amplified DNAs F1/R1, F2/R3, and F4/R4 amplified DNAs and was designated library Y5.

10

#### EXAMPLE 12

##### Immunoscreening of the Y5 Library

###### A. ISOLATION OF IMMUNOREACTIVE CLONES.

Two HGV positive sera, PNF2161 and JC, were used for immunoscreening of the Y5 library, essentially as 15 described in Example 2. The Y5 phage library was plated onto 20 plates at approximately 15,000 phage per plate. The plates were incubated for approximately 5 hours and were overlaid with nitrocellulose filters (Schleicher and Schuell) overnight. The filters were blocked by 20 incubation in AIB (1% gelatin plus 0.02% Na azide) for approximately 6 hours. The blocked filters were washed once with TBS.

Ten Y5 library filters were incubated overnight, with agitation, with PNF2161 serum and ten filters with JC 25 serum. In order to reduce non-specific antibody binding, both sera had been pre-treated by incubation overnight with nitrocellulose filters to which wild type λgt11 were adsorbed.

The filters were removed from the sera, washed 3 30 times with TBS and incubated with goat anti-human alkaline phosphatase-conjugated secondary antibody (Promega; diluted 1/7500 in AIB) for one hour. The filters were washed 4 times with TBS. Bound secondary antibody was detected by incubation of the filters in AP buffer (100 mM 35 NaCl, 5 mM MgCl<sub>2</sub>, 100 mM Tris pH 9.5) containing NBT & BCIP.

Plaques that tested positive in the initial screen were picked and eluted in 500  $\mu$ l of PDB (100 mM NaCl, 8.1 mM MgSO<sub>4</sub>, 50 mM Tris pH 7.5, 0.02% Gelatin). The immunoreactive phage were purified by replating the eluted 5 phage at a total density of 100 - 500 plaques per 100 mm plate. The plates were re-immunoscreened with the appropriate HGV-positive sera, essentially as described above. After color development several isolated, positive plaques were picked and put into 500  $\mu$ l of PDB. After 1 10 hour of incubation, 2  $\mu$ l of the re-purified phage PDB solution was used as template in a PCR reaction containing the 11F (SEQ ID NO:25) and 11R (SEQ ID NO:13) PCR primers. These primers are homologous to sequences located 70 15 nucleotides (nt) 5' and 90 nt 3' of the EcoRI site of  $\lambda$ gt11. The PCR reactions were amplified through 30 cycles of 94°C for 1 minute, 55°C for 1.5 minutes and 72°C for 2 minutes.

The PCR amplification reactions were size-fractionated on agarose gels. PCR amplification of 20 purified plaques resulted in a single band for each single-plaque amplification reaction, where the amplified fragment contained the DNA insert plus approximately 140 bp of 5' and 3' phage flanking sequences. The amplified products, from PCR reactions resulting in single bands, 25 were purified using a "S-300 HR" spin column (Pharmacia), following manufacturers instructions. The DNA was quantitated and DNA sequenced employing an Applied Biosystems automated sequencer 373A and appropriate protocols.

30 The above-described screening of the Y5 library with JC sera resulted in the purification and DNA sequencing of the positive-strand clones presented in Table 11. Positive-strand clones correspond to the 5' to 3' translation of the HGV sequence presented in SEQ ID NO:14 35 -- the polyprotein reading frame.

Table 11

<b>Clone</b>	<b>Screening Sera</b>	<b>Insert Size (base pairs)</b>	<b>Insert Size (amino acids)</b>	<b>Nucleic Acid SEQ ID NO:</b>	<b>Encoded Protein SEQ ID NO:</b>
5	Y5-10	JC	210	62	48
	Y5-12	JC	333	94	50
	Y5-26	JC	303	93	52
	Y5-5	JC	153	36	54
	Y5-3	JC	162	44	56
10	Y5-27	JC	288	86	58
	Y5-25	JC	165	36	60
	Y5-20	JC	165	19 <sup>1</sup>	62
	Y5-16	JC	234	63	64
15	the clone contained a double insert, nt 69 to 126 of the clone insert correspond to HGV sequences.				

These clones delineated 2 immunogenic regions within the putative NS5 protein of HGV. These two regions are specifically delineated by Y5-10 and Y5-5.

20 Further, screening of the Y5 library with PNF 2161 sera resulted in the purification and DNA sequencing of the following negative-strand clones presented in Table 12. Negative-strand clones correspond to the 5' to 3' translation of the sequence complementary to the HGV 25 sequence presented in SEQ ID NO:14.

Table 12

<b>Clone</b>	<b>Screening Sera</b>	<b>Insert Size (base pairs)</b>	<b>Insert Size (amino acids)</b>	<b>Nucleic Acid SEQ ID NO:</b>	<b>Encoded Protein SEQ ID NO:</b>
30	Y5-50	PNF 2161	349	104	66
	Y5-52	PNF 2161	119	20 <sup>1</sup>	68
	Y5-53	PNF 2161	250	33 <sup>2</sup>	70
	Y5-55	PNF 2161	143	19 <sup>3</sup>	72

Clone	Screening Sera	Insert Size (base pairs)	Insert Size (amino acids)	Nucleic Acid SEQ ID NO:	Encoded Protein SEQ ID NO:
Y5-56	PNF 2161	366	110	74	75
Y5-57	PNF 2161	231	65	76	77
Y5-60	PNF 2161	151	38	78	79
Y5-63	PNF 2161	125 <sup>4</sup>	25	80	81

- 5      1      the clone contained a double insert, nt 46 to 105 of  
the clone insert correspond to HGV sequences.
- 10     2      the clone contained a double insert, nt 19 to 118 of  
the clone insert correspond to HGV sequences.
- 15     3      the clone contained a double insert, nt 70 to 126 of  
the clone insert correspond to HGV sequences.
- 15     4      the insert contains an extra, non-HGV sequence  
between nucleotides 19 and 35.

20      All of these sequences contain portions of the  
original HGV clone 470-20-1 isolated using the PNF 2161  
serum.

#### B. FURTHER CHARACTERIZATION OF IMMUNOREACTIVE CLONES.

25      Clones Y5-10, Y5-16, and Y5-5 were selected for sub-  
cloning into the expression vector pGEX-HisB. PCR primers  
were designed which removed the extraneous linker  
sequences at the end of these clones. These primers also  
introduced (i) a *Nco*I site at the 5' end (relative to the  
coding sequence) of each insert, and (ii) a *Bam*HI site at  
30     the 3' end of each insert. Using these primers (see Table  
13), the DNA fragments were amplified from 2  $\mu$ l of the  
plaque pure stocks.

Table 13

Clone	Primer Set		
Y5-10	Y5-10-F1 Y5-10-R1	SEQ ID NO:82 SEQ ID NO:83	
Y5-16	Y5-16F1 470ep-R3	SEQ ID NO:84 SEQ ID NO:85	
5	Y5-5	Y5-5-F1 470ep-R3	SEQ ID NO:86 SEQ ID NO:85

- Amplifications were performed as follows: 30 cycles of 94°C for 1 minute, 50°C for 1.5 minutes, and 72°C for 2 minutes. After amplification the resulting DNAs were 10 purified using "WIZARD PCR," spin columns, the samples eluted in 50 µl, and digested overnight with *NcoI* and *BamHI*. A minimum of 30 units of each enzyme was used in the restriction endonuclease digestions (*NcoI*, Boehringer Mannheim; *BamHI*, Promega).
- 15 The digested PCR fragments were ligated overnight to expression vector pGEX-HisB that had been digested with *NcoI* and *BamHI*. Each set of ligated plasmids was independently used to transform *E. coli* strain W3110, using a heat shock protocol (Ausubel, et al.; Maniatis, et 20 al.). Transformants were selected on LB plates containing 100 µg/ml ampicillin and resistant colonies were used to inoculate 2 mls of LB containing 100 µg/ml ampicillin. Cultures expressing non-recombinant sj26/his protein were also prepared.
- 25 After incubation overnight at 37°C the cultures were diluted 1/10 into 2 mls of fresh LB plus ampicillin and grown for an additional 1 hour at 37°C. IPTG was added to a final concentration of 0.2 mM and the cultures were grown for an additional 3 hours at 37°C. The bacteria 30 were pelleted by centrifugation and the bacterial pellet was resuspended in 100 µl PBS. To the pellet, 100 µl of

2X SDS sample buffer (0.125 M Tris, pH 6.8, 10% glycine, 5%  $\beta$ -mercaptoethanol, 2.3% SDS) was added. The resulting lysates were vortexed and heated to 100°C for 5 minutes. Aliquots (15  $\mu$ l) of each lysate were loaded onto a 12% acrylamide SDS-PAGE gel.

The expressed proteins were size-fractionated by electrophoresis. The separated proteins were transferred from the gel to nitrocellulose filters using standard techniques (Harlow, et al.). An additional gel containing 10 the expressed proteins was stained using coomasie blue protein stain.

Transformants carrying plasmids Y5-10, Y5-5 and Y5-16 expressed significant amounts of correctly sized recombinant fusion proteins. The identity of the 15 recombinant fusions were confirmed by incubating a Western blot (prepared above) with a murine monoclonal antibody that is specifically immunoreactive with sj26 (Sierra BioSource, Gilroy, CA).

Additional confirmation that the picked colonies 20 contained the appropriate insert was obtained as follows. A phage solution for each colony was prepared by inoculating 40  $\mu$ l of TE solution with a toothpick containing a small amount of bacteria putatively expressing a recombinant clone had been inoculated. A 5 25  $\mu$ l sample was taken from each solution and separately PCR amplified.

The amplifications employed the appropriate forward primer, (e.g., Y5-10 F for a colony putatively expressing Y5-10) and a reverse primer (SEQ ID NO:87) homologous to a 30 sequence located 3' to the cloning sites of the plasmid pGEX-HisB. The PCR amplifications were for 25 cycles as follows: 94°C for 1 minute, 50°C for 1.5 minutes and 72°C for 2 minutes. All of the colonies selected for further analysis produced a correctly sized DNA band with no other 35 obvious bands under these conditions.

The immunoreactivity of the antigens expressed from the Y5-10, Y5-16, & Y5-5 inserts (expressed as sj26-his

fusion proteins) was determined as follows. Aliquots (15 µl) of the crude lysates prepared above were size-fractionated by SDS-PAGE using a 12% acrylamide gel. The proteins were electro-blotted ("NOVEX MINICELL MINIBLOT II," San Diego, CA) onto nitrocellulose filters. The filters were then individually incubated with one of the following sera: JC, PNF 2161, and Super normal serum 4 (SN4) (R05072) as a negative control. In addition, one filter was incubated with anti-sj26 monoclonal antibodies (RM001; Sierra BioSource).

As expected, the recombinant protein produced by the bacteria expressing the antigens encoded by the Y5-10, Y5-5, and Y5-16 inserts all reacted with JC sera. No reactivity was observed with either PNF 2161 or SN4 sera. All proteins appeared to be expressed at similar levels as determined by their reactivity to the anti-sj26 monoclonal antibody. The Y5-5 and Y5-10 encoded proteins were selected for further purification.

*E. coli* carrying Y5-5- and Y5-10- containing pGEX-20 HisB vectors were cultured and expression of the fusion protein induced as described above. The cells were lysed in PBS, containing 2 mM PMSF, using a French Press at 1500 psi. The crude lysate was spun to remove cellular debris. The supernatant was loaded onto the glutathione affinity column at a high flow rate and the column was washed with 10 column volumes of PBS. The Y5-5 and Y5-10 fusion proteins were eluted with 10 mM Tris pH 8.8 containing 10 mM glutathione.

Each of the fusion protein samples was diluted 1/10 with Buffer A (10 mM Tris pH 8.8, containing 8 M urea) and loaded onto a nickel charged-chelating "SEPHAROSE" fast flow column. Each column was repeatedly washed with Buffer A until no further contaminants were eluted. The fusion proteins were eluted using a gradient of imidazole in buffer A. An imidazole gradient was run from 0 to 0.5 M imidazole in 20 column volumes. Fractions were collected.

Each set of fractions was analyzed by standard SDS-PAGE using 12% polyacrylamide gels. Pools of the Y5-5 and Y5-10 fusion protein-containing fractions were separately made.

5 Figures 8A to 8D show the results of Western blot analysis of the following samples ( $\mu$ g/lane): lane 1, Y5-10 antigen 1.6  $\mu$ g; lane 2, Y5-10 antigen 0.8  $\mu$ g; lane 3, Y5-10 antigen 0.4  $\mu$ g; and lane 4, Y5-10 antigen 0.2  $\mu$ g. Human serum JC (Figure 8A) and Super Normal 2 serum  
10 (Figure 8B) were diluted 1:100. The anti-GST mouse monoclonal antibody RM001 (Figure 8C) was diluted 1:1000. Figure 8D shows the Y5-10 antigen resolved by SDS-PAGE, transferred onto the nitrocellulose membrane and stained with Ponceau S protein stain (Kodak, Rochester, NY;  
15 Sigma). Arrow indicates the location of Y5.10 antigen. These results demonstrate that Y5-10 is specifically immunoreactive with N-(ABCDE) human serum JC.

Figures 9A to 9D show the results of Western blot analysis of the following samples: lane 1, Y5-5 antigen  
20 3.2  $\mu$ g; lane 2, Y5-5 antigen 1.6  $\mu$ g; lane 3, Y5-5 antigen 0.8  $\mu$ g; lane 4, Y5-5 antigen 0.4  $\mu$ g; lane 5, Y5-5 antigen 0.2  $\mu$ g; lane 6, GE3-2 antigen 0.4  $\mu$ g; and lane 7, Y5-10 antigen 0.4  $\mu$ g. Human serum JC (Figure 9A), T55806 (Figure 9B), and Super Normal 2 serum (Figure 9C) were  
25 diluted 1:100. RM001, the anti-GST mouse monoclonal antibody, (Figure 9D) was diluted 1:1000. Arrows indicate the locations of antigens Y5.5, GE3.2 and Y5.10. These results show specific immunoreactivity of the Y5-5 antigen with the JC serum. Further, the antigens GE3-2 and Y5-10  
30 were reactive with T55806. However, the Y5-5 antigen was not reactive with the HGV-positive sera T55806.

The Y5-10 antigen was also size-fractionated by SDS polyacrylamide gel electrophoresis. The gel was stained using coomasie blue protein stain. The gel was scanned for purity with a laser densitometer. The purity of the Y5-10 fusion protein was approximately 95%.

EXAMPLE 13Cloning Further HGV Isolates

## A. THE JC VARIANT.

One milliliter of JC serum was spun at 40,000 rpms  
5 for 2 hours. The resulting pellet was extracted using "TRIREAGENT" (MRC, Cincinnati, OH), resulting in the formation of 3 phases. The upper phase contained RNA only. This phase was taken and ETOH precipitated.

HGV cDNA molecules were generated from the JC sample  
10 by two methods. The first method was amplification (RT-PCR) of the JC nucleic acid sample using specific and nested primers. The primer sequences were based on the HGV sequence obtained from PNF 2161 serum. The criteria used to select the primers were (i) regions having a high  
15 G/C content, and (ii) no repetitious sequences.

The second method used to generate HGV cDNA molecules was amplification using HGV (PNF 2161) specific primers followed by identification of HGV specific sequences with <sup>32</sup>P-labelled oligonucleotide probes. Such DNA  
20 hybridizations were carried out essentially as described by Sambrook, et al. (1989). The PCR derived clones were either (i) cloned into the "TA" vector (Invitrogen, San Diego, CA) and sequenced with vector primers (TAR and TAF), or (ii) sequenced directly after PCR amplification.  
25 Both the probe and primer sequences were based on the HGV variant obtained from the PNF 2161 serum.

These two approaches yielded multiply-overlapping HGV fragments from the JC serum. Each of these fragments were cloned and sequenced. The sequences were aligned to  
30 obtain the HGV (JC-variant) consensus sequence presented as SEQ ID NO:156 (polypeptide sequence, SEQ ID NO:157). The sequence of each region of the HGV (JC-variant) virus was based on a consensus from at least three different, overlapping, independent clones.

**B. OTHER HGV VARIANTS.**

In addition to the HGV PNF 2161-variant and JC-variant sequences, three partial HGV isolates have been obtained from the sera BG34, T55806 and EB20 by methods similar to those described above. The partial sequences of these isolates are presented as SEQ ID NO:150 (BG34 nucleic acid), SEQ ID NO:151 (BG34 protein), SEQ ID NO:152 (T55806 nucleic acid), SEQ ID NO:153 (T55806 protein), SEQ ID NO:154 (EB20-2 nucleic acid) and SEQ ID NO:155 (EB20-2 protein).

While the invention has been described with reference to specific methods and embodiments, it will be appreciated that various modifications and changes may be made without departing from the invention.

110

SEQUENCE LISTING

## (1) GENERAL INFORMATION:

## (i) APPLICANT:

- (A) NAME: Genelabs Technologies, Inc.
- (B) STREET: 505 Penobscot Drive
- (C) CITY: Redwood City
- (D) STATE: CA
- (E) COUNTRY: USA
- (F) POSTAL CODE: 94063

(ii) TITLE OF INVENTION: Detection of Viral Antigens Coded  
by Reverse Reading Frames

(iii) NUMBER OF SEQUENCES: 157

## (iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Dehlinger & Associates
- (B) STREET: 350 Cambridge Avenue, Suite 250
- (C) CITY: Palo Alto
- (D) STATE: CA
- (E) COUNTRY: USA
- (F) ZIP: 94306

## (v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25

## (vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER:
- (B) FILING DATE:
- (C) CLASSIFICATION:

## (vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/246,985
- (B) FILING DATE: 20-MAY-1994

## (viii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/285,561
- (B) FILING DATE: 03-AUG-1994

## (ix) PRIOR APPLICATION DATA:

111

- (A) APPLICATION NUMBER: US 08/329,729
- (B) FILING DATE: 26-OCT-1994

## (vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/344,271
- (B) FILING DATE: 23-NOV-1994

## (viii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/357,509
- (B) FILING DATE: 16-DEC-1994

## (ix) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/389,886
- (B) FILING DATE: 15-FEB-1995

## (viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Fabian, Gary R.
- (B) REGISTRATION NUMBER: 33,875
- (C) REFERENCE/DOCKET NUMBER: 4600-0202.41

## (ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (415) 324-0880
- (B) TELEFAX: (415) 324-0960

## (2) INFORMATION FOR SEQ ID NO:1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (iii) HYPOTHETICAL: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: SISPA primer, top strand Linker AB

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GGAATTCGCG GCCGCTCG

112

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Linker AB, bottom strand

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CGAGCGGCCG CGAATTCCCTT

20

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 237 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: PNF 2161 CLONE 470-20-1

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..237

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GAA TTC GCG GCC GCT CGG GCT GTC TCG GAC TCT TGG ATG ACC TCG AAT

48

113

Glu Phe Ala Ala Ala Arg Ala Val Ser Asp Ser Trp Met Thr Ser Asn				
1	5	10	15	
GAG TCA GAG GAC GGG GTA TCC TCC TGC GAG GAG GAC ACC GGC GGG GTC				96
Glu Ser Glu Asp Gly Val Ser Ser Cys Glu Asp Thr Gly Gly Val				
20	25	30		
TTC TCA TCT GAG CTG CTC TCA GTA ACC GAG ATA AGT GCT GGC GAT GGA				144
Phe Ser Ser Glu Leu Leu Ser Val Thr Glu Ile Ser Ala Gly Asp Gly				
35	40	45		
GTA CGG GGG ATG TCT TCT CCC CAT ACA ACA GGC ATC TCT CGG CTA CTA CCA				192
Val Arg Gly Met Ser Ser Pro His Thr Gly Ile Ser Arg Leu Leu Pro				
50	55	60		
CAA AGA GAG GGT GTA CTG CAG TCC TCC ACG AGC GGC CGC GAA TTC				237
Gln Arg Glu Gly Val Leu Gln Ser Ser Thr Ser Gly Arg Glu Phe				
65	70	75		

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 79 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Glu Phe Ala Ala Ala Arg Ala Val Ser Asp Ser Trp Met Thr Ser Asn			
1	5	10	15
Glu Ser Glu Asp Gly Val Ser Ser Cys Glu Asp Thr Gly Gly Val			
20	25	30	
Phe Ser Ser Glu Leu Leu Ser Val Thr Glu Ile Ser Ala Gly Asp Gly			
35	40	45	
Val Arg Gly Met Ser Ser Pro His Thr Gly Ile Ser Arg Leu Leu Pro			
50	55	60	
Gln Arg Glu Gly Val Leu Gln Ser Ser Thr Ser Gly Arg Glu Phe			

114

65

70

75

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: HAV-R1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GTTGACCAAC TGAGTCTGAA GC

22

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: HAV-F1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GATTGGAAAT CTGATCCGTC CC

22

115

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: HCV-LANR

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TCGCGACCCA ACACTACTC

19

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: HCV 1532

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GGGGCGACA CTCCACCA

18

## (2) INFORMATION FOR SEQ ID NO:9:

116

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer 470-20-1-77F

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CTCTTTGTGG TAGTAGCCGA GAGAT

25

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 24 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer 470-20-1-211R

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CGAATGAGTC AGAGGACGGG GTAT

24

117

## (2) INFORMATION FOR SEQ ID NO:11:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Primer KL-1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GCAGGATCCG AATTGGCATC TAGAGAT

27

## (2) INFORMATION FOR SEQ ID NO:12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Primer KL-2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

ATCTCTAGAT GCGAATTCCG ATCCTGCGA

29

## (2) INFORMATION FOR SEQ ID NO:13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

## (vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: LAMBDA GT11, REVERSE PRIMER

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GGCAGACATG GCCTGCCCGG

20

## (2) INFORMATION FOR SEQ ID NO:14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9391 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: HGV-PNF 2161 Variant

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 459..9077

119

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

ACGTGGGGGA GTTGATCCCC CCCCCCGGC ACTGGGTGCA AGCCCCAGAA ACCGACGCCT	60		
ATCTAAGTAG ACCGCAATGAC TCGGCGCCGA CTCGGCGACC GGCCAAAAGG TGGTGGATGG	120		
GTGATGACAG GGTTGGTAGG TCGTAAATCC CGGTACCTT GGTAGCCACT ATAGGTGGGT	180		
CTTAAGAGAA GGTTAAGATT CCTCTTGTGC CTGCGCGAG ACCGCGCACG GTCCACAGGT	240		
GTTGGCCCTA CCCGTGGGAA TAAGGGCCCG ACGTCAGGCT CGTCGTTAAA CCGAGCCCGT	300		
TACCCACCTG GGCAAACGAC GCCCACGTAC GGTCCACGTC GCCCTTCAAT GTCTCTCTTG	360		
ACCAATAGGC TAGCCGGCG AGTTGACAAG GACCAGTGGG GGCCGGGGGC TTGGAGAGGG	420		
ACTCCAAGTC CGGCCCTTCC CGGTGGGCCG GGAAATGC ATG GGG CCA CCC AGC	473		
Met Gly Pro Pro Ser			
1	5		
TCC GCG GCG GCC TGC AGC CGG GGT AGC CCA AGA ATC CTT CGG GTG AGG	521		
Ser Ala Ala Ala Cys Ser Arg Gly Ser Pro Arg Ile Leu Arg Val Arg			
10	15	20	
GGC GGT GGC ATT TCC TTT TTC TAT ACC ATC ATG GCA GTC CTT CTG CTC	569		
Ala Gly Gly Ile Ser Phe Phe Tyr Thr Ile Met Ala Val Leu Leu Leu			
25	30	35	
CTT CTC GTG GTT GAG GCC GGG GCC ATT CTG GCC CCG GCC ACC CAC GCT	617		
Leu Leu Val Val Glu Ala Gly Ala Ile Leu Ala Pro Ala Thr His Ala			
40	45	50	
TGT CGA GCG AAT GGG CAA TAT TTC CTC ACA AAT TGT TGT GCC CCG GAG	665		
Cys Arg Ala Asn Gly Gln Tyr Phe Leu Thr Asn Cys Cys Ala Pro Glu			
55	60	65	
GAC ATC GGG TTC TGC CTG GAG GGT GGA TGC CTG GTG GCC CTG GGG TGC	713		
Asp Ile Gly Phe Cys Leu Glu Gly Gly Cys Leu Val Ala Leu Gly Cys			
70	75	80	85
ACG ATT TGC ACT GAC CAA TGC TGG CCA CTG TAT CAG GCG GGT TTG GCT	761		
Thr Ile Cys Thr Asp Gln Cys Trp Pro Leu Tyr Gln Ala Gly Leu Ala			
90	95	100	

			120	
GTG CGG CCT GGC AAG TCC GCG GCC CAA CTG GTG GGG GAG CTG GGT AGC Val Arg Pro Gly Lys Ser Ala Ala Gln Leu Val Gly Glu Leu Gly Ser	105	110	115	809
CTA TAC GGG CCC CTG TCG GTC TCG GCC TAT GTG GCT GGG ATC CTG GGC Leu Tyr Gly Pro Leu Ser Val Ser Ala Tyr Val Ala Gly Ile Leu Gly	120	125	130	857
CTG GGT GAG GTG TAC TCG GGT GTC CTA ACG GTG GGA GTC GCG TTG ACG Leu Gly Glu Val Tyr Ser Gly Val Leu Thr Val Gly Val Ala Leu Thr	135	140	145	905
CGC CGG GTC TAC CCG GTG CCT AAC CTG ACG TGT GCA GTC GCG TGT GAG Arg Arg Val Tyr Pro Val Pro Asn Leu Thr Cys Ala Val Ala Cys Glu	150	155	160	953
CTA AAG TGG GAA AGT GAG TTT TGG AGA TGG ACT GAA CAG CTG GCC TCC Leu Lys Trp Glu Ser Glu Phe Trp Arg Trp Thr Glu Gln Leu Ala Ser	170	175	180	1001
AAC TAC TGG ATT CTG GAA TAC CTC TGG AAG GTC CCA TTT GAT TTC TGG Asn Tyr Trp Ile Leu Glu Tyr Leu Trp Lys Val Pro Phe Asp Phe Trp	185	190	195	1049
AGA GGC GTG ATA AGC CTG ACC CCC TTG TTG GTT TGC GTG GCC GCA TTG Arg Gly Val Ile Ser Leu Thr Pro Leu Leu Val Cys Val Ala Ala Leu	200	205	210	1097
CTG CTG CTT GAG CAA CGG ATT GTC ATG GTC TTC CTG TTG GTG ACG ATG Leu Leu Leu Glu Gln Arg Ile Val Met Val Phe Leu Leu Val Thr Met	215	220	225	1145
GCC GGG ATG TCG CAA GGC GCC CCT GCC TCC GTT TTG GGG TCA CGC CCC Ala Gly Met Ser Gln Gly Ala Pro Ala Ser Val Leu Gly Ser Arg Pro	230	235	240	1193
TTT GAC TAC GGG TTG ACT TGG CAG ACC TGC TCT TGC AGG GCC AAC GGT Phe Asp Tyr Gly Leu Thr Trp Gln Thr Cys Ser Cys Arg Ala Asn Gly	250	255	260	1241
TCG CGT TTT TCG ACT GGG GAG AAG GTG TGG GAC CGT GGG AAC GTT ACG Ser Arg Phe Ser Thr Gly Glu Lys Val Trp Asp Arg Gly Asn Val Thr	265	270	275	1289

121

CTT CAG TGT GAC TGC CCT AAC GGC CCC TGG GTG TGG TTG CCA GCC TTT Leu Gln Cys Asp Cys Pro Asn Gly Pro Trp Val Trp Leu Pro Ala Phe 280                   285                   290	1337
TGC CAA GCA ATC GGC TGG GGT GAC CCC ATC ACT TAT TGG AGC CAC GGG Cys Gln Ala Ile Gly Trp Gly Asp Pro Ile Thr Tyr Trp Ser His Gly 295                   300                   305	1385
CAA AAT CAG TGG CCC CTT TCA TGC CCC CAG TAT GTC TAT GGG TCT GCT Gln Asn Gln Trp Pro Leu Ser Cys Pro Gln Tyr Val Tyr Gly Ser Ala 310                   315                   320                   325	1433
ACA GTC ACT TGC GTG TGG GGT TCC GCT TCT TGG TTT GCC TCC ACC AGT Thr Val Thr Cys Val Trp Gly Ser Ala Ser Trp Phe Ala Ser Thr Ser 330                   335                   340	1481
GGT CGC GAC TCG AAG ATA GAT GTG TGG AGT TTA GTG CCA GTT GGC TCT Gly Arg Asp Ser Lys Ile Asp Val Trp Ser Leu Val Pro Val Gly Ser 345                   350                   355	1529
GCC ACC TGC ACC ATA GCC GCA CTT GGA TCA TCG GAT CGC GAC ACG GTG Ala Thr Cys Thr Ile Ala Ala Leu Gly Ser Ser Asp Arg Asp Thr Val 360                   365                   370	1577
CCT GGG CTC TCC GAG TGG GGA ATC CCG TGC GTG ACG TGT GTT CTG GAC Pro Gly Leu Ser Glu Trp Gly Ile Pro Cys Val Thr Cys Val Leu Asp 375                   380                   385	1625
CGT CGG CCT GCC TCC TGC CGC ACC TGT GTG AGG GAC TGC TGG CCC GAG Arg Arg Pro Ala Ser Cys Gly Thr Cys Val Arg Asp Cys Trp Pro Glu 390                   395                   400                   405	1673
ACC GGG TCG GTT AGG TTC CCA TTC CAT CGG TGC GGC GTG GGG CCT CGG Thr Gly Ser Val Arg Phe Pro Phe His Arg Cys Gly Val Gly Pro Arg 410                   415                   420	1721
CTG ACA AAG GAC TTG GAA GCT GTG CCC TTC GTC AAC AGG ACA ACT CCC Leu Thr Lys Asp Leu Glu Ala Val Pro Phe Val Asn Arg Thr Thr Pro 425                   430                   435	1769
TTC ACC ATT AGG GGG CCC CTG GGC AAC CAG GGC CGA GGC AAC CCG GTG Phe Thr Ile Arg Gly Pro Leu Gly Asn Gln Gly Arg Gly Asn Pro Val 440                   445                   450	1817

122

CGG TCG CCC TTG GGT TTT GGG TCC TAC GCC ATG ACC AGC ATC CGA GAT Arg Ser Pro Leu Gly Phe Gly Ser Tyr Ala Met Thr Arg Ile Arg Asp	455	460	465	1865
ACC CTA CAT CTG GTG GAG TGT CCC ACA CCA GCC ATT GAG CCT CCC ACC Thr Leu His Leu Val Glu Cys Pro Thr Pro Ala Ile Glu Pro Pro Thr	470	475	480	485
Gly Thr Phe Gly Phe Pro Gly Thr Pro Pro Leu Asn Asn Cys Met	490	495	500	1913
GGG ACG TTT GGG TTC TTC CCC GGG ACG CCG CCT CTC AAC AAC TGC ATG Gly Thr Phe Gly Phe Pro Gly Thr Pro Pro Leu Asn Asn Cys Met	505	510	515	1961
CTC TTG GGC ACG GAA GTG TCC GAG GCA CTT GGG GGG GCT GGC CTC ACG Leu Leu Gly Thr Glu Val Ser Glu Ala Leu Gly Gly Ala Gly Leu Thr	520	525	530	2009
Gly Gly TTC TAT GAA CCC CTG GTG CGC AGG TGT TCG AAG CTG ATG GGA Gly Phe Tyr Glu Pro Leu Val Arg Arg Cys Ser Lys Leu Met Gly	535	540	545	2057
AGC CGA AAT CCG GTT TGT CCG GGG TTT GCA TGG CTC TCT TCG GGC AGG Ser Arg Asn Pro Val Cys Pro Gly Phe Ala Trp Leu Ser Ser Gly Arg	550	555	560	2105
CCT GAT GGG TTT ATA CAT GTC CAG GGT CAC TTG CAG GAG GTG GAT GCA Pro Asp Gly Phe Ile His Val Gln Gly His Leu Gln Glu Val Asp Ala	565	570	575	2153
GCC AAC TTC ATC CCG CCC CCG CGC TGG TTG CTC TTG GAC TTT GTA TTT Gly Asn Phe Ile Pro Pro Arg Trp Leu Leu Asp Phe Val Phe	580	585	590	2201
GTC CTG TTA TAC CTG ATG AAG CTG GCT GAG GCA CGG TTG GTC CCG CTG Val Leu Leu Tyr Leu Met Lys Leu Ala Glu Ala Arg Leu Val Pro Leu	595	600	605	2249
ATC TTG CTG CTG CTA TGG TGG TGG GTG AAC CAG CTG GCA GTC CTA GGG Ile Leu Leu Leu Trp Trp Trp Val Asn Gln Leu Ala Val Leu Gly	610	615	620	2297
CTG CCG GCT GTG GAA GCC GCC GTG GCA GGT GAG GTC TTC GCG GGC CCT Leu Pro Ala Val Glu Ala Ala Val Ala Gly Glu Val Phe Ala Gly Pro	625	630	635	2345

123

GCC CTG TCC TGG TGT CTG GGA CTC CCG GTC GTC AGT ATG ATA TTG GGT Ala Leu Ser Trp Cys Leu Gly Leu Pro Val Val Ser Met Ile Leu Gly 630 635 640 645	2393
TTG GCA AAC CTG GTG CTG TAC TTT AGA TGG TTG GGA CCC CAA CGC CTG Leu Ala Asn Leu Val Leu Tyr Phe Arg Trp Leu Gly Pro Gln Arg Leu 650 655 660	2441
ATG TTC CTC GTG TTG TGG AAG CTT GCT CGG GGA GCT TTC CCG CTG GCC Met Phe Leu Val Leu Trp Lys Leu Ala Arg Gly Ala Phe Pro Leu Ala 665 670 675	2489
CTC TTG ATG GGG ATT TCG GCG ACC CGC GGG CGC ACC TCA GTG CTC GGG Leu Leu Met Gly Ile Ser Ala Thr Arg Gly Arg Thr Ser Val Leu Gly 680 685 690	2537
GCC GAG TTC TGC TTC GAT GCT ACA TTC GAG GTG GAC ACT TCG GTG TTG Ala Glu Phe Cys Phe Asp Ala Thr Phe Glu Val Asp Thr Ser Val Leu 695 700 705	2585
GGC TGG GTG GTG GCC AGT GTG GTA GCT TGG GCC ATT GCG CTC CTG AGC Gly Trp Val Val Ala Ser Val Val Ala Trp Ala Ile Ala Leu Leu Ser 710 715 720 725	2633
TCG ATG AGC GCA GGG GGG TGG AGG CAC AAA GCC GTG ATC TAT AGG ACG Ser Met Ser Ala Gly Gly Trp Arg His Lys Ala Val Ile Tyr Arg Thr 730 735 740	2681
TGG TGT AAG GGG TAC CAG GCA ATC CGT CAA AGG GTG GTG AGG AGC CCC Trp Cys Lys Gly Tyr Gln Ala Ile Arg Gln Arg Val Val Arg Ser Pro 745 750 755	2729
CTC GGG GAG GGG CGG CCT GCC AAA CCC CTG ACC TTT GCC TGG TGC TTG Leu Gly Glu Gly Arg Pro Ala Lys Pro Leu Thr Phe Ala Trp Cys Leu 760 765 770	2777
GCC TCG TAC ATC TGG CCA GAT GCT GTG ATG ATG GTG GTG GTT GCC TTG Ala Ser Tyr Ile Trp Pro Asp Ala Val Met Met Val Val Val Ala Leu 775 780 785	2825
GTC CTT CTC TTT GGC CTG TTC GAC GCG TTG GAT TGG GCC TTG GAG GAG Val Leu Leu Phe Gly Leu Phe Asp Ala Leu Asp Trp Ala Leu Glu Glu 790 795 800 805	2873

124

ATC TTG GTG TCC CCG CCC TCG TTG CGG CGT TTG GCT CGG GTG GTT GAG Ile Leu Val Ser Arg Pro Ser Leu Arg Arg Leu Ala Arg Val Val Glu 810	815	820	2921
TGC TGT GTG ATG GCG GGT GAG AAG GCC ACA ACC GTC CGG CTG GTC TCC Cys Cys Val Met Ala Gly Glu Lys Ala Thr Thr Val Arg Leu Val Ser 825	830	835	2969
AAG ATG TGT GCG AGA GGA GCT TAT TTG TTC GAT CAT ATG GGC TCT TTT Lys Met Cys Ala Arg Gly Ala Tyr Leu Phe Asp His Met Gly Ser Phe 840	845	850	3017
TCG CGT GCT GTC AAC GAG CGC CTG TTG GAA TGG GAC GCA GCT CTT GAA Ser Arg Ala Val Lys Glu Arg Leu Leu Glu Trp Asp Ala Ala Leu Glu 855	860	865	3065
CCT CTG TCA TTC ACT AGG ACG GAC TGT CGC ATC ATA CGG GAT GCC GCG Pro Leu Ser Phe Thr Arg Thr Asp Cys Arg Ile Ile Arg Asp Ala Ala 870	875	880	885
AGG ACT TTG TCC TGC GGG CAG TGC GTC ATG GGT TTA CCC GTG GTT GCG Arg Thr Leu Ser Cys Gly Gln Cys Val Met Gly Leu Pro Val Val Ala 890	895	900	3113
CGC CGT GGT GAT GAG GTT CTC ATC GGC GTC TTC CAG GAT GTG AAT CAT Arg Arg Gly Asp Glu Val Leu Ile Gly Val Phe Gln Asp Val Asn His 905	910	915	3161
TTG CCT CCC GGG TTT GTT CCG ACC GCG CCT GTT GTC ATC CGA CGG TGC Leu Pro Pro Gly Phe Val Pro Thr Ala Pro Val Val Ile Arg Arg Cys 920	925	930	3209
GGA AAG GGC TTC TTG GGG GTC ACA AAG GCT GCC TTG ACA GGT CGG GAT Gly Lys Gly Phe Leu Gly Val Thr Lys Ala Ala Leu Thr Gly Arg Asp 935	940	945	3257
CCT GAC TTA CAT CCA GGG AAC GTC ATG GTG TTG GGG ACG GCT ACG TCG Pro Asp Leu His Pro Gly Asn Val Met Val Leu Gly Thr Ala Thr Ser 950	955	960	3305
CGA AGC ATG GGA ACA TGC TTG AAC GGC CTG CTG TTC ACG ACC TTC CAT Arg Ser Met Gly Thr Cys Leu Asn Gly Leu Leu Phe Thr Thr Phe His 970	975	980	3353
			3401

125

GGG GCT TCA TCC CGA ACC ATC GCC ACA CCC GTG GGG GCC CTT AAT CCC Gly Ala Ser Ser Arg Thr Ile Ala Thr Pro Val Gly Ala Leu Asn Pro 985	990	995	3449
AGA TGG TGG TCA GCC AGT GAT GAT GTC ACG GTG TAT CCA CTC CCG GAT Arg Trp Trp Ser Ala Ser Asp Asp Val Thr Val Tyr Pro Leu Pro Asp 1000	1005	1010	3497
GGG GCT ACT TCG TTA ACA CCT TGT ACT TGC CAG GCT GAG TCC TGT TGG Gly Ala Thr Ser Leu Thr Pro Cys Thr Cys Gln Ala Glu Ser Cys Trp 1015	1020	1025	3545
GTC ATC AGA TCC GAC GGG GCC CTA TGC CAT GGC TTG AGC AAG GGG GAC Val Ile Arg Ser Asp Gly Ala Leu Cys His Gly Leu Ser Lys Gly Asp 1030	1035	1040	1045
AAG GTG GAG CTG GAT GTG GCC ATG GAG GTC TCT GAC TTC CGT GGC TCG Lys Val Glu Leu Asp Val Ala Met Glu Val Ser Asp Phe Arg Gly Ser 1050	1055	1060	3641
TCT GGC TCA CCG GTC CTA TGT GAC GAA GGG CAC GCA GTC GGA ATG CTC Ser Gly Ser Pro Val Leu Cys Asp Glu Gly His Ala Val Gly Met Leu 1065	1070	1075	3689
GTG TCT GTG CTT CAC TCC GGT GGT AGG GTC ACC GCG GCA CGG TTC ACT Val Ser Val Leu His Ser Gly Gly Arg Val Thr Ala Ala Arg Phe Thr 1080	1085	1090	3737
AGG CCG TGG ACC CAA GTG CCA ACA GAT GCC AAA ACC ACT ACT GAA CCC Arg Pro Trp Thr Gln Val Pro Thr Asp Ala Lys Thr Thr Glu Pro 1095	1100	1105	3785
CCT CCG GTG CCG GCC AAA GGA GTT TTC AAA GAG GGC CCG TTG TTT ATG Pro Pro Val Pro Ala Lys Gly Val Phe Lys Glu Ala Pro Leu Phe Met 1110	1115	1120	3833
CCT ACG GGA GCG GGA AAG AGC ACT CGC GTC CCG TTG GAG TAC GAT AAC Pro Thr Gly Ala Gly Lys Ser Thr Arg Val Pro Leu Glu Tyr Asp Asn 1130	1135	1140	3881
ATG GGG CAC AAG GTC TTA ATC TTG AAC CCC TCA GTG GCC ACT GTG CGG Met Gly His Lys Val Leu Ile Leu Asn Pro Ser Val Ala Thr Val Arg 1145	1150	1155	3929

126

GCC ATG GGC CCG TAC ATG GAG CGG CTG GCC GGT AAA CAT CCA AGT ATA Ala Met Gly Pro Tyr Met Glu Arg Leu Ala Gly Lys His Pro Ser Ile 1160                   1165                   1170	3977
TAC TGT GGG CAT GAT ACA ACT GCT TTC ACA AGG ATC ACT GAC TCC CCC Tyr Cys Gly His Asp Thr Thr Ala Phe Thr Arg Ile Thr Asp Ser Pro 1175                   1180                   1185	4025
CTG ACG TAT TCA ACC TAT GGG AGG TTT TTG GCC AAC CCT AGG CAG ATG Leu Thr Tyr Ser Thr Tyr Gly Arg Phe Leu Ala Asn Pro Arg Gln Met 1190                   1195                   1200                   1205	4073
CTA CGG GGC GTT TCG GTG GTC ATT TGT GAT GAG TGC CAC AGT CAT GAC Leu Arg Gly Val Ser Val Val Ile Cys Asp Glu Cys His Ser His Asp 1210                   1215                   1220	4121
TCA ACC GTG CTG TTA GGC ATT GGG AGA GTC CGG GAG CTG GCG CGT GGG Ser Thr Val Leu Leu Gly Ile Gly Arg Val Arg Glu Leu Ala Arg Gly 1225                   1230                   1235	4169
TGC GGG GTG CAA CTA GTG CTC TAC GCC ACC GCT ACA CCT CCC GGA TCC Cys Gly Val Gln Leu Val Leu Tyr Ala Thr Ala Thr Pro Pro Gly Ser 1240                   1245                   1250	4217
CCT ATG ACG CAG CAC CCT TCC ATA ATT GAG ACA AAA TTG GAC GTG GGC Pro Met Thr Gln His Pro Ser Ile Ile Glu Thr Lys Leu Asp Val Gly 1255                   1260                   1265	4265
GAG ATT CCC TTT TAT GGG CAT GGA ATA CCC CTC GAG CGG ATG CGA ACC Glu Ile Pro Phe Tyr Gly His Gly Ile Pro Leu Glu Arg Met Arg Thr 1270                   1275                   1280                   1285	4313
GGA AGG CAC CTC GTG TTC TGC CAT TCT AAG GCT GAG TGC GAG CGC CTT Gly Arg His Leu Val Phe Cys His Ser Lys Ala Glu Cys Glu Arg Leu 1290                   1295                   1300	4361
GCT GGC CAG TTC TCC GCT AGG GGG GTC AAT GCC ATT GCC TAT TAT AGG Ala Gly Gln Phe Ser Ala Arg Gly Val Asn Ala Ile Ala Tyr Tyr Arg 1305                   1310                   1315	4409
GGT AAA GAC AGT TCT ATC ATC AAG GAT GGG GAC CTG GTG GTC TGT GCT Gly Lys Asp Ser Ser Ile Ile Lys Asp Gly Asp Leu Val Val Cys Ala 1320                   1325                   1330	4457

127

ACA GAC GCG CTT TCC ACT GGG TAC ACT GGA AAT TTC GAC TCC GTC ACC			4505
Thr Asp Ala Leu Ser Thr Gly Tyr Thr Gly Asn Phe Asp Ser Val Thr			
1335	1340	1345	
GAC TGT GGA TTA GTG GTG GAG GAG GTC GTT GAG GTG ACC CTT GAT CCC			4553
Asp Cys Gly Leu Val Val Glu Glu Val Val Glu Val Thr Leu Asp Pro			
1350	1355	1360	1365
ACC ATT ACC ATC TCC CTG CGG ACA GTG CCT GCG TCG GCT GAA CTG TCG			4601
Thr Ile Thr Ile Ser Leu Arg Thr Val Pro Ala Ser Ala Glu Leu Ser			
1370	1375	1380	
ATG CAA AGA CGA GGA CGC ACG GGT AGG GGC AGG TCT GGA CGC TAC TAC			4649
Met Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Ser Gly Arg Tyr Tyr			
1385	1390	1395	
TAC GCG GGG GTG GGC AAA GCC CCT GCG GGT GTG GTG CGC TCA GGT CCT			4697
Tyr Ala Gly Val Gly Lys Ala Pro Ala Gly Val Val Arg Ser Gly Pro			
1400	1405	1410	
GTC TGG TCG GCG GTG GAA GCT GGA GTG ACC TGG TAC GGA ATG GAA CCT			4745
Val Trp Ser Ala Val Glu Ala Gly Val Thr Trp Tyr Gly Met Glu Pro			
1415	1420	1425	
GAC TTG ACA GCT AAC CTA CTG AGA CTT TAC GAC GAC TGC CCT TAC ACC			4793
Asp Leu Thr Ala Asn Leu Leu Arg Leu Tyr Asp Asp Cys Pro Tyr Thr			
1430	1435	1440	1445
GCA GCC GTC GCG GCT GAT ATC GGA GAA GCC GCG GTG TTC TTC TCT GGG			4841
Ala Ala Val Ala Ala Asp Ile Gly Glu Ala Ala Val Phe Phe Ser Gly			
1450	1455	1460	
CTC GCC CCA TTG AGG ATG CAC CCT GAT GTC AGC TGG GCA AAA GTT CGC			4889
Leu Ala Pro Leu Arg Met His Pro Asp Val Ser Trp Ala Lys Val Arg			
1465	1470	1475	
GGC GTC AAC TGG CCC CTC TTG GTG GGT GTT CAG CGG ACC ATG TGT CGG			4937
Gly Val Asn Trp Pro Leu Leu Val Gly Val Gln Arg Thr Met Cys Arg			
1480	1485	1490	
GAA ACA CTG TCT CCC GGC CCA TCG GAT GAC CCC CAA TGG GCA GGT CTG			4985
Glu Thr Leu Ser Pro Gly Pro Ser Asp Asp Pro Gln Trp Ala Gly Leu			
1495	1500	1505	

128

AAG	GGC	CCA	AAT	CCT	GTC	CCA	CTC	CTG	CTG	AGG	TGG	GGC	AAT	GAT	TTA	5033
Lys	Gly	Pro	Asn	Pro	Val	Pro	Leu	Leu	Leu	Arg	Trp	Gly	Asn	Asp	Leu	
1510				1515				1520							1525	
CCA	TCT	AAA	GTG	GCC	GGC	CAC	CAC	ATA	GTG	GAC	GAC	CTG	GTC	CGG	AGA	5081
Pro	Ser	Lys	Val	Ala	Gly	His	His	Ile	Val	Asp	Asp	Leu	Val	Arg	Arg	
			1530					1535							1540	
CTC	GGT	GTG	GCG	GAG	GGT	TAC	GTC	CGC	TGC	GAC	GCT	GGG	CCG	ATC	TTG	5129
Leu	Gly	Val	Ala	Glu	Gly	Tyr	Val	Arg	Cys	Asp	Ala	Gly	Pro	Ile	Leu	
			1545				1550								1555	
ATG	ATC	GGT	CTA	GCT	ATC	GCG	GGG	GGG	ATG	ATC	TAC	GCG	TCA	TAC	ACC	5177
Met	Ile	Gly	Leu	Ala	Ile	Ala	Gly	Gly	Met	Ile	Tyr	Ala	Ser	Tyr	Thr	
			1560			1565				1570						
GGG	TCG	CTA	GTG	GTG	GTG	ACA	GAC	TGG	GAT	GTG	AAG	GGG	GGT	GGC	GCC	5225
Gly	Ser	Leu	Val	Val	Val	Thr	Asp	Trp	Asp	Val	Lys	Gly	Gly	Gly	Ala	
			1575			1580				1585						
CCC	CTT	TAT	CGG	CAT	GGA	GAC	CAG	GCC	ACG	CCT	CAG	CCG	GTG	GTG	CAG	5273
Pro	Leu	Tyr	Arg	His	Gly	Asp	Gln	Ala	Thr	Pro	Gln	Pro	Val	Val	Gln	
			1590			1595			1600						1605	
GTT	CCT	CCG	GTA	GAC	CAT	CGG	CCG	GGG	GGT	GAA	TCA	GCA	CCA	TCG	GAT	5321
Val	Pro	Pro	Val	Asp	His	Arg	Pro	Gly	Gly	Glu	Ser	Ala	Pro	Ser	Asp	
			1610			1615				1620						
GCC	AAG	ACA	GTG	ACA	GAT	GCG	GTG	GCA	GCC	ATC	CAG	GTG	GAC	TGC	GAT	5369
Ala	Lys	Thr	Val	Thr	Asp	Ala	Val	Ala	Ala	Ile	Gln	Val	Asp	Cys	Asp	
			1625			1630				1635						
TGG	ACT	ATC	ATG	ACT	CTG	TCG	ATC	GGA	GAA	GTG	TTG	TCC	TTG	GCT	CAG	5417
Trp	Thr	Ile	Met	Thr	Leu	Ser	Ile	Gly	Glu	Val	Leu	Ser	Leu	Ala	Gln	
			1640			1645				1650						
GCT	AAG	ACG	GCC	GAG	GCC	TAC	ACA	GCA	ACC	GCC	AAG	TGG	CTC	GCT	GGC	5465
Ala	Lys	Thr	Ala	Glu	Ala	Tyr	Thr	Ala	Thr	Ala	Lys	Trp	Leu	Ala	Gly	
			1655			1660				1665						
TGC	TAT	ACG	GGG	ACG	CGG	GCC	GTT	CCC	ACT	GTA	TCC	ATT	GTT	GAC	AAG	5513
Cys	Tyr	Thr	Gly	Thr	Arg	Ala	Val	Pro	Thr	Val	Ser	Ile	Val	Asp	Lys	
			1670			1675				1680					1685	

129

CTC TTC GCC GGA GGG TGG GCG GCT GTG GTG GGC CAT TGC CAC AGC GTG		5561
Leu Phe Ala Gly Gly Trp Ala Ala Val Val Gly His Cys His Ser Val		
1690	1695	1700
ATT GCT GCG GCG GTG GCG GCC TAC GGG GCT TCA AGG AGC CCG CCG TTG		5609
Ile Ala Ala Ala Val Ala Ala Tyr Gly Ala Ser Arg Ser Pro Pro Leu		
1705	1710	1715
GCA GCC GCG GCT TCC TAC CTG ATG GGG TTG GGC GTT GGA GGC AAC GCT		5657
Ala Ala Ala Ala Ser Tyr Leu Met Gly Leu Gly Val Gly Gly Asn Ala		
1720	1725	1730
CAG ACG CGC CTG GCG TCT GCC CTC CTA TTG GGG GCT GCT GGA ACC GCC		5705
Gln Thr Arg Leu Ala Ser Ala Leu Leu Leu Gly Ala Ala Gly Thr Ala		
1735	1740	1745
TTG GGC ACT CCT GTC GTG GGC TTG ACC ATG GCA GGT GCG TTC ATG GGG		5753
Leu Gly Thr Pro Val Val Gly Leu Thr Met Ala Gly Ala Phe Met Gly		
1750	1755	1760
1765		
GGG GCC AGT GTC TCC CCC TCC TTG GTC ACC ATT TTA TTG GGG GCC GTC		5801
Gly Ala Ser Val Ser Pro Ser Leu Val Thr Ile Leu Leu Gly Ala Val		
1770	1775	1780
GGA GGT TGG GAG GGT GTT GTC AAC GCG GCG AGC CTA GTC TTT GAC TTC		5849
Gly Gly Trp Glu Gly Val Val Asn Ala Ala Ser Leu Val Phe Asp Phe		
1785	1790	1795
ATG GCG GGG AAA CTT TCA TCA GAA GAT CTG TGG TAT GCC ATC CCG GTA		5897
Met Ala Gly Lys Leu Ser Ser Glu Asp Leu Trp Tyr Ala Ile Pro Val		
1800	1805	1810
CTG ACC AGC CCG GGG GCG CTT GCG GGG ATC GCT CTC GGG TTG GTT		5945
Leu Thr Ser Pro Gly Ala Gly Leu Ala Gly Ile Ala Leu Gly Leu Val		
1815	1820	1825
TTG TAT TCA GCT AAC AAC TCT GGC ACT ACC ACT TGG TTG AAC CGT CTG		5993
Leu Tyr Ser Ala Asn Asn Ser Gly Thr Thr Thr Trp Leu Asn Arg Leu		
1830	1835	1840
1845		
CTG ACT ACG TTA CCA AGG TCT TCA TGT ATC CCG GAC AGT TAC TTT CAG		6041
Leu Thr Thr Leu Pro Arg Ser Ser Cys Ile Pro Asp Ser Tyr Phe Gln		
1850	1855	1860

130

CAA GTT GAC TAT TGC GAC AAG GTC TCA GCC GTG CTC CGG CGC CTG AGC Gln Val Asp Tyr Cys Asp Lys Val Ser Ala Val Leu Arg Arg Leu Ser 1865	1870	1875	6089
CTC ACC CGC ACA GTG GTT GCC CTG GTC AAC AGG GAG CCT AAG GTG GAT Leu Thr Arg Thr Val Val Ala Leu Val Asn Arg Glu Pro Lys Val Asp 1880	1885	1890	6137
GAG GTA CAG GTG GGG TAT GTC TGG GAC CTG TGG GAG TGG ATC ATG CGC Glu Val Gln Val Gly Tyr Val Trp Asp Leu Trp Glu Trp Ile Met Arg 1895	1900	1905	6185
CAA GTG CGC GTG GTC ATG GCC AGA CTC AGG GCC CTC TGC CCC GTG GTG Gln Val Arg Val Val Met Ala Arg Leu Arg Ala Leu Cys Pro Val Val 1910	1915	1920	1925
TCA CTA CCC TTG TGG CAT TGC GGG GAG GGG TGG TCC GGG GAA TGG TTG Ser Leu Pro Leu Trp His Cys Gly Glu Gly Trp Ser Gly Glu Trp Leu 1930	1935	1940	6281
CTT GAC GGT CAT GTT GAG AGT CGC TGC CTC TGT GGC TGC GTG ATC ACT Leu Asp Gly His Val Glu Ser Arg Cys Leu Cys Gly Cys Val Ile Thr 1945	1950	1955	6329
GGT GAC GTT CTG AAT GGG CAA CTC AAA GAA CCA GTT TAC TCT ACC AAG Gly Asp Val Leu Asn Gly Gln Leu Lys Glu Pro Val Tyr Ser Thr Lys 1960	1965	1970	6377
CTG TGC CGG CAC TAT TGG ATG GGG ACT GTC CCT GTG AAC ATG CTG GGT Leu Cys Arg His Tyr Trp Met Gly Thr Val Pro Val Asn Met Leu Gly 1975	1980	1985	6425
TAC GGT GAA ACG TCG CCT CTC CTG GCC TCC GAC ACC CCG AAG GTT GTG Tyr Gly Glu Thr Ser Pro Leu Leu Ala Ser Asp Thr Pro Lys Val Val 1990	1995	2000	2005
CCC TTC GGG ACG TCT GGC TGG GCT GAG GTG GTG GTG ACC ACT ACC CAC Pro Phe Gly Thr Ser Gly Trp Ala Glu Val Val Val Thr Thr His 2010	2015	2020	6473
GTG GTA ATC AGG AGG ACC TCC GCC TAT AAG CTG CTG CGC CAG CAA ATC Val Val Ile Arg Arg Thr Ser Ala Tyr Lys Leu Leu Arg Gln Gln Ile 2025	2030	2035	6521
			6569

131

CTA TCG GCT GCT GTA GCT GAG CCC TAC TAC GTC GAC GGC ATT CCG GTC		6617
Leu Ser Ala Ala Val Ala Glu Pro Tyr Tyr Val Asp Gly Ile Pro Val		
2040	2045	2050
TCA TGG GAC GCG GAC GCT CGT GCG CCC GCC ATG GTC TAT GGC CCT GGG		6665
Ser Trp Asp Ala Asp Ala Arg Ala Pro Ala Met Val Tyr Gly Pro Gly		
2055	2060	2065
CAA AGT GTT ACC ATT GAC GGG GAG CGC TAC ACC TTG CCT CAT CAA CTG		6713
Gln Ser Val Thr Ile Asp Gly Glu Arg Tyr Thr Leu Pro His Gln Leu		
2070	2075	2080
2085		
AGG CTC AGG AAT GTG GCA CCC TCT GAG GTT TCA TCC GAG GTG TCC ATT		6761
Arg Leu Arg Asn Val Ala Pro Ser Glu Val Ser Ser Glu Val Ser Ile		
2090	2095	2100
GAC ATT GGG ACG GAG ACT GAA GAC TCA GAA CTG ACT GAG GCC GAT CTG		6809
Asp Ile Gly Thr Glu Thr Glu Asp Ser Glu Leu Thr Glu Ala Asp Leu		
2105	2110	2115
CCG CCG GCG GCT GCT CTC CAA GCG ATC GAG AAT GCT GCG AGG ATT		6857
Pro Pro Ala Ala Ala Leu Gln Ala Ile Glu Asn Ala Ala Arg Ile		
2120	2125	2130
CTT GAA CCG CAC ATT GAT GTC ATC ATG GAG GAC TGC AGT ACA CCC TCT		6905
Leu Glu Pro His Ile Asp Val Ile Met Glu Asp Cys Ser Thr Pro Ser		
2135	2140	2145
CTT TGT GGT AGT AGC CGA GAG ATG CCT GTA TGG GGA GAA GAC ATC CCC		6953
Leu Cys Gly Ser Ser Arg Glu Met Pro Val Trp Gly Glu Asp Ile Pro		
2150	2155	2160
2165		
CGT ACT CCA TCG CCA GCA CTT ATC TCG GTT ACT GAG ACC AGC TCA GAT		7001
Arg Thr Pro Ser Pro Ala Leu Ile Ser Val Thr Glu Ser Ser Asp		
2170	2175	2180
GAG AAG ACC CCG TCG GTG TCC TCC TCG CAG GAG GAT ACC CCG TCC TCT		7049
Glu Lys Thr Pro Ser Val Ser Ser Gln Glu Asp Thr Pro Ser Ser		
2185	2190	2195
GAC TCA TTC GAG GTC ATC CAA GAG TCC GAG ACA GCA GGG GAG GAA		7097
Asp Ser Phe Glu Val Ile Gln Glu Ser Glu Thr Ala Glu Gly Glu Glu		
2200	2205	2210

132

AGT GTC TTC AAC GTG GCT CTT TCC GTA TTA AAA GCC TTA TTT CCA CAG Ser Val Phe Asn Val Ala Leu Ser Val Leu Lys Ala Leu Phe Pro Gln	2215	2220	2225	7145
AGC GAC GCG ACC AGG AAG CTT ACC GTC AAG ATG TCG TGC TGC GTT GAA Ser Asp Ala Thr Arg Lys Leu Thr Val Lys Met Ser Cys Cys Val Glu	2230	2235	2240	2245
AAG AGC GTC ACG CGC TTT TTC TCA TTG GGG TTG ACG GTG GCT GAT GTT Lys Ser Val Thr Arg Phe Phe Ser Leu Gly Leu Thr Val Ala Asp Val	2250	2255	2260	7241
GCT AGC CTG TGT GAG ATG GAA ATC CAG AAC CAT ACA GCC TAT TGT GAC Ala Ser Leu Cys Glu Met Glu Ile Gln Asn His Thr Ala Tyr Cys Asp	2265	2270	2275	7289
CAG GTG CGC ACT CCG CTT GAA TTG CAG GTT GGG TGC TTG GTG GGC AAT Gln Val Arg Thr Pro Leu Glu Leu Gln Val Gly Cys Leu Val Gly Asn	2280	2285	2290	7337
GAA CTT ACC TTT GAA TGT GAC AAG TGT GAG GCT AGG CAA GAA ACC TTG Glu Leu Thr Phe Glu Cys Asp Lys Cys Glu Ala Arg Gln Glu Thr Leu	2295	2300	2305	7385
GCC TCC TTC TCT TAC ATT TGG TCT GGA GTG CCG CTG ACT AGG GCC ACC Ala Ser Phe Ser Tyr Ile Trp Ser Gly Val Pro Leu Thr Arg Ala Thr	2310	2315	2320	2325
CCG GCC AAG CCT CCC GTG GTG AGG CCG GTT GGC TCT TTG TTA GTG GCC Pro Ala Lys Pro Pro Val Val Arg Pro Val Gly Ser Leu Leu Val Ala	2330	2335	2340	7481
GAC ACT ACT AAG GTG TAT GTT ACC AAT CCA GAC AAT GTG GGA CGG AGG Asp Thr Thr Lys Val Tyr Val Thr Asn Pro Asp Asn Val Gly Arg Arg	2345	2350	2355	7529
GTG GAC AAG GTG ACC TTC TGG CGT GCT CCT AGG GTT CAT GAT AAG TAC Val Asp Lys Val Thr Phe Trp Arg Ala Pro Arg Val His Asp Lys Tyr	2360	2365	2370	7577
CTC GTG GAC TCT ATT GAG CGC GCT AAG AGG GCC GCT CAA GCC TGC CTA Leu Val Asp Ser Ile Glu Arg Ala Lys Arg Ala Ala Gln Ala Cys Leu	2375	2380	2385	7625

133

AGC ATG GGT TAC ACT TAT GAG GAA GCA ATA AGG ACT GTA AGG CCA CAT Ser Met Gly Tyr Thr Tyr Glu Glu Ala Ile Arg Thr Val Arg Pro His	7673
2390                    2395                    2400                    2405	
GCT GCC ATG GGC TGG GGA TCT AAG GTG TCG GTT AAG GAC TTA GCC ACC Ala Ala Met Gly Trp Gly Ser Lys Val Ser Val Lys Asp Leu Ala Thr	7721
2410                    2415                    2420	
CCC GCG GGG AAG ATG GCC GTC CAT GAC CGG CTT CAG GAG ATA CTT GAA Pro Ala Gly Lys Met Ala Val His Asp Arg Leu Gln Glu Ile Leu Glu	7769
2425                    2430                    2435	
GGG ACT CCG GTC CCC TTT ACT CTT ACT GTG AAA AAG GAG GTG TTC TTC Gly Thr Pro Val Pro Phe Thr Leu Thr Val Lys Lys Glu Val Phe Phe	7817
2440                    2445                    2450	
AAA GAC CGG AAG GAG GAG AAG GCC CCC CGC CTC ATT GTG TTC CCC CCC Lys Asp Arg Lys Glu Glu Lys Ala Pro Arg Leu Ile Val Phe Pro Pro	7865
2455                    2460                    2465	
CTG GAC TTC CGG ATA GCT GAA AAG CTC ATC TTG GGA GAC CCA GGC CGG Leu Asp Phe Arg Ile Ala Glu Lys Leu Ile Leu Gly Asp Pro Gly Arg	7913
2470                    2475                    2480                    2485	
GTA GCC AAG GCG GTG TTG GGG GGG GCC TAC GCC TTC CAG TAC ACC CCA Val Ala Lys Ala Val Leu Gly Gly Ala Tyr Ala Phe Gln Tyr Thr Pro	7961
2490                    2495                    2500	
AAT CAG CGA GTT AAG GAG ATG CTC AAG CTA TGG GAG TCT AAG AAG ACC Asn Gln Arg Val Lys Glu Met Leu Lys Leu Trp Glu Ser Lys Lys Thr	8009
2505                    2510                    2515	
CCT TGC GCC ATC TGT GTG GAC GCC ACC TGC TTC GAC AGT AGC ATA ACT Pro Cys Ala Ile Cys Val Asp Ala Thr Cys Phe Asp Ser Ser Ile Thr	8057
2520                    2525                    2530	
GAA GAG GAC GTG GCT TTG GAG ACA GAG CTA TAC GCT CTG GCC TCT GAC Glu Glu Asp Val Ala Leu Glu Thr Glu Leu Tyr Ala Leu Ala Ser Asp	8105
2535                    2540                    2545	
CAT CCA GAA TGG GTG CGG GCA CTT GGG AAA TAC TAT GCC TCA GGC ACC His Pro Glu Trp Val Arg Ala Leu Gly Lys Tyr Tyr Ala Ser Gly Thr	8153
2550                    2555                    2560                    2565	

134

ATG GTC ACC CCG GAA GGG GTG CCC GTC GGT GAG AGG TAT TGC AGA TCC Met Val Thr Pro Glu Gly Val Pro Val Gly Glu Arg Tyr Cys Arg Ser	2570	2575	2580	8201
TCG GGT GTC CTA ACA ACT AGC GCG AGC AAC TGC TTG ACC TGC TAC ATC Ser Gly Val Leu Thr Thr Ser Ala Ser Asn Cys Leu Thr Cys Tyr Ile	2585	2590	2595	8249
AAG GTG AAA GCT GCC TGT GAG AGA GTG GGG CTG AAA AAT GTC TCT CTT Lys Val Lys Ala Ala Cys Glu Arg Val Gly Leu Lys Asn Val Ser Leu	2600	2605	2610	8297
CTC ATA GCC GGC GAT GAC TGC TTG ATC ATA TGT GAG CGG CCA GTG TGC Leu Ile Ala Gly Asp Asp Cys Leu Ile Ile Cys Glu Arg Pro Val Cys	2615	2620	2625	8345
GAC CCA AGC GAC GCT TTG GGC AGA GCC CTA CGG AGC TAT GGG TAC GCG Asp Pro Ser Asp Ala Leu Gly Arg Ala Leu Ala Ser Tyr Gly Tyr Ala	2630	2635	2640	2645
TGC GAG CCC TCA TAT CAT GCA TCA TTG GAC ACG GCC CCC TTC TGC TCC Cys Glu Pro Ser Tyr His Ala Ser Leu Asp Thr Ala Pro Phe Cys Ser	2650	2655	2660	8441
ACT TGG CTT GCT GAG TGC AAT GCA GAT GGG AAG CGC CAT TTC TTC CTG Thr Trp Leu Ala Glu Cys Asn Ala Asp Gly Lys Arg His Phe Phe Leu	2665	2670	2675	8489
ACC ACG GAC TTC CCG AGG CCG CTC GCT CGC ATG TCG AGT GAG TAT ACT Thr Thr Asp Phe Arg Arg Pro Leu Ala Arg Met Ser Ser Glu Tyr Ser	2680	2685	2690	8537
GAC CCG ATG GCT TCG GCG ATC GGT TAC ATC CTC CTT TAT CCT TGG CAC Asp Pro Met Ala Ser Ala Ile Gly Tyr Ile Leu Leu Tyr Pro Trp His	2695	2700	2705	8585
CCC ATC ACA CGG TGG GTC ATC ATC CCT CAT GTG CTA ACG TGC GCA TTC Pro Ile Thr Arg Trp Val Ile Ile Pro His Val Leu Thr Cys Ala Phe	2710	2715	2720	2725
AGG GGT GGA GGC ACA CCG TCT GAT CCG GTT TGG TGC CAG GTG CAT GGT Arg Gly Gly Gly Thr Pro Ser Asp Pro Val Trp Cys Gln Val His Gly	2730	2735	2740	8681

135

AAC TAC TAC AAG TTT CCA CTG GAC AAA CTG CCT AAC ATC ATC GTG GCC Asn Tyr Tyr Lys Phe Pro Leu Asp Lys Leu Pro Asn Ile Ile Val Ala 2745 2750 2755	8729
CTC CAC GGA CCA GCA GCG TTG AGG GTT ACC GCA GAC ACA ACT AAA ACA Leu His Gly Pro Ala Ala Leu Arg Val Thr Ala Asp Thr Thr Lys Thr 2760 2765 2770	8777
AAG ATG GAG GCT GGT AAG GTT CTG AGC GAC CTC AAG CTC CCT GGC TTA Lys Met Glu Ala Gly Lys Val Leu Ser Asp Leu Lys Leu Pro Gly Leu 2775 2780 2785	8825
GCA GTC CAC CGA AAG AAG GCC GGG GCG TTG CGA ACA CGC ATG CTC CGC Ala Val His Arg Lys Lys Ala Gly Ala Leu Arg Thr Arg Met Leu Arg 2790 2795 2800 2805	8873
TCG CGC GGT TGG GCT GAG TTG GCT AGG GGC TTG TTG TGG CAT CCA GGC Ser Arg Gly Trp Ala Glu Leu Ala Arg Gly Leu Leu Trp His Pro Gly 2810 2815 2820	8921
CTA CGG CTT CCT CCC CCT GAG ATT GCT GGT ATC CCG GGG GGT TTC CCT Leu Arg Leu Pro Pro Pro Glu Ile Ala Gly Ile Pro Gly Gly Phe Pro 2825 2830 2835	8969
CTC TCC CCC CCC TAT ATG GGG GTG GTA CAT CAA TTG GAT TTC ACA AGC Leu Ser Pro Pro Tyr Met Gly Val Val His Gln Leu Asp Phe Thr Ser 2840 2845 2850	9017
CAG AGG AGT CGC TGG CGG TGG TTG GGG TTC TTA GCC CTG CTC ATC GTA Gln Arg Ser Arg Trp Arg Trp Leu Gly Phe Leu Ala Leu Ile Val 2855 2860 2865	9065
GCC CTC TTC GGG TGAACCTAAAT TCATCTGTTG CGGCAAGGTC TGGTACTGA Ala Leu Phe Gly 2870	9117
TCATCACCGG AGGAGGTTCC CGCCCTCCCC GCCCCAGGGG TCTCCCCGCT GGGTAAAAAG	9177
GGCCCGGCCT TGGGAGGCAT GGTGGTTACT AACCCCCTGG CAGGGTCAA GCCTGATGGT	9237
GCTAATGCAC TGCCACTTCG GTGGCGGGTC GCTACCTTAT AGCGTAATCC GTGACTACGG	9297
GCTGCTCGCA GAGCCCTCCC CGGATGGGGC ACAGTGCAGT GTGATCTGAA GGGGTGCACC	9357

136

CCGGGAAGAG CTCGGCCCGA AGGCCGGTTC TACT

9391

## (2) INFORMATION FOR SEQ ID NO:15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2873 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Met Gly Pro Pro Ser Ser Ala Ala Ala Cys Ser Arg Gly Ser Pro Arg  
1 5 10 15

Ile Leu Arg Val Arg Ala Gly Gly Ile Ser Phe Phe Tyr Thr Ile Met  
20 25 30

Ala Val Leu Leu Leu Leu Val Val Glu Ala Gly Ala Ile Leu Ala  
35 40 45

Pro Ala Thr His Ala Cys Arg Ala Asn Gly Gln Tyr Phe Leu Thr Asn  
50 55 60

Cys Cys Ala Pro Glu Asp Ile Gly Phe Cys Leu Glu Gly Gly Cys Leu  
65 70 75 80

Val Ala Leu Gly Cys Thr Ile Cys Thr Asp Gln Cys Trp Pro Leu Tyr  
85 90 95

Gln Ala Gly Leu Ala Val Arg Pro Gly Lys Ser Ala Ala Gln Leu Val  
100 105 110

Gly Glu Leu Gly Ser Leu Tyr Gly Pro Leu Ser Val Ser Ala Tyr Val  
115 120 125

Ala Gly Ile Leu Gly Leu Gly Glu Val Tyr Ser Gly Val Leu Thr Val  
130 135 140

Gly Val Ala Leu Thr Arg Arg Val Tyr Pro Val Pro Asn Leu Thr Cys  
145 150 155 160

137

Ala Val Ala Cys Glu Leu Lys Trp Glu Ser Glu Phe Trp Arg Trp Thr  
165 170 175

Glu Gln Leu Ala Ser Asn Tyr Trp Ile Leu Glu Tyr Leu Trp Lys Val  
180 185 190

Pro Phe Asp Phe Trp Arg Gly Val Ile Ser Leu Thr Pro Leu Leu Val  
195 200 205

Cys Val Ala Ala Leu Leu Leu Glu Gln Arg Ile Val Met Val Phe  
210 215 220

Leu Leu Val Thr Met Ala Gly Met Ser Gln Gly Ala Pro Ala Ser Val  
225 230 235 240

Leu Gly Ser Arg Pro Phe Asp Tyr Gly Leu Thr Trp Gln Thr Cys Ser  
245 250 255

Cys Arg Ala Asn Gly Ser Arg Phe Ser Thr Gly Glu Lys Val Trp Asp  
260 265 270

Arg Gly Asn Val Thr Leu Gln Cys Asp Cys Pro Asn Gly Pro Trp Val  
275 280 285

Trp Leu Pro Ala Phe Cys Gln Ala Ile Gly Trp Gly Asp Pro Ile Thr  
290 295 300

Tyr Trp Ser His Gly Gln Asn Gln Trp Pro Leu Ser Cys Pro Gln Tyr  
305 310 315 320

Val Tyr Gly Ser Ala Thr Val Thr Cys Val Trp Gly Ser Ala Ser Trp  
325 330 335

Phe Ala Ser Thr Ser Gly Arg Asp Ser Lys Ile Asp Val Trp Ser Leu  
340 345 350

Val Pro Val Gly Ser Ala Thr Cys Thr Ile Ala Ala Leu Gly Ser Ser  
355 360 365

Asp Arg Asp Thr Val Pro Gly Leu Ser Glu Trp Gly Ile Pro Cys Val  
370 375 380

Thr Cys Val Leu Asp Arg Arg Pro Ala Ser Cys Gly Thr Cys Val Arg  
385 390 395 400

138

Asp Cys Trp Pr Glu Thr Gly Ser Val Arg Phe Pro Phe His Arg Cys  
405 410 415

Gly Val Gly Pro Arg Leu Thr Lys Asp Leu Glu Ala Val Pro Phe Val  
420 425 430

Asn Arg Thr Thr Pro Phe Thr Ile Arg Gly Pro Leu Gly Asn Gln Gly  
435 440 445

Arg Gly Asn Pro Val Arg Ser Pro Leu Gly Phe Gly Ser Tyr Ala Met  
450 455 460

Thr Arg Ile Arg Asp Thr Leu His Leu Val Glu Cys Pro Thr Pro Ala  
465 470 475 480

Ile Glu Pro Pro Thr Gly Thr Phe Gly Phe Pro Gly Thr Pro Pro  
485 490 495

Leu Asn Asn Cys Met Leu Leu Gly Thr Glu Val Ser Glu Ala Leu Gly  
500 505 510

Gly Ala Gly Leu Thr Gly Gly Phe Tyr Glu Pro Leu Val Arg Arg Cys  
515 520 525

Ser Lys Leu Met Gly Ser Arg Asn Pro Val Cys Pro Gly Phe Ala Trp  
530 535 540

Leu Ser Ser Gly Arg Pro Asp Gly Phe Ile His Val Gln Gly His Leu  
545 550 555 560

Gln Glu Val Asp Ala Gly Asn Phe Ile Pro Pro Pro Arg Trp Leu Leu  
565 570 575

Leu Asp Phe Val Phe Val Leu Leu Tyr Leu Met Lys Leu Ala Glu Ala  
580 585 590

Arg Leu Val Pro Leu Ile Leu Leu Leu Trp Trp Trp Val Asn Gln  
595 600 605

Leu Ala Val Leu Gly Leu Pro Ala Val Glu Ala Ala Val Ala Gly Glu  
610 615 620

Val Phe Ala Gly Pro Ala Leu Ser Trp Cys Leu Gly Leu Pro Val Val  
625 630 635 640

139

Ser Met Ile Leu Gly Leu Ala Asn Leu Val Leu Tyr Phe Arg Trp Leu  
645 650 655

Gly Pro Gln Arg Leu Met Phe Leu Val Leu Trp Lys Leu Ala Arg Gly  
660 665 670

Ala Phe Pro Leu Ala Leu Leu Met Gly Ile Ser Ala Thr Arg Gly Arg  
675 680 685

Thr Ser Val Leu Gly Ala Glu Phe Cys Phe Asp Ala Thr Phe Glu Val  
690 695 700

Asp Thr Ser Val Leu Gly Trp Val Val Ala Ser Val Val Ala Trp Ala  
705 710 715 720

Ile Ala Leu Leu Ser Ser Met Ser Ala Gly Gly Trp Arg His Lys Ala  
725 730 735

Val Ile Tyr Arg Thr Trp Cys Lys Gly Tyr Gln Ala Ile Arg Gln Arg  
740 745 750

Val Val Arg Ser Pro Leu Gly Glu Gly Arg Pro Ala Lys Pro Leu Thr  
755 760 765

Phe Ala Trp Cys Leu Ala Ser Tyr Ile Trp Pro Asp Ala Val Met Met  
770 775 780

Val Val Val Ala Leu Val Leu Leu Phe Gly Leu Phe Asp Ala Leu Asp  
785 790 795 800

Trp Ala Leu Glu Glu Ile Leu Val Ser Arg Pro Ser Leu Arg Arg Leu  
805 810 815

Ala Arg Val Val Glu Cys Cys Val Met Ala Gly Glu Lys Ala Thr Thr  
820 825 830

Val Arg Leu Val Ser Lys Met Cys Ala Arg Gly Ala Tyr Leu Phe Asp  
835 840 845

His Met Gly Ser Phe Ser Arg Ala Val Lys Glu Arg Leu Leu Glu Trp  
850 855 860

Asp Ala Ala Leu Glu Pro Leu Ser Phe Thr Arg Thr Asp Cys Arg Ile  
865 870 875 880

140

Ile Arg Asp Ala Ala Arg Thr Leu Ser Cys Gly Gln Cys Val Met Gly  
885 890 895

Leu Pro Val Val Ala Arg Arg Gly Asp Glu Val Leu Ile Gly Val Phe  
900 905 910

Gln Asp Val Asn His Leu Pro Pro Gly Phe Val Pro Thr Ala Pro Val  
915 920 925

Val Ile Arg Arg Cys Gly Lys Gly Phe Leu Gly Val Thr Lys Ala Ala  
930 935 940

Leu Thr Gly Arg Asp Pro Asp Leu His Pro Gly Asn Val Met Val Leu  
945 950 955 960

Gly Thr Ala Thr Ser Arg Ser Met Gly Thr Cys Leu Asn Gly Leu Leu  
965 970 975

Phe Thr Thr Phe His Gly Ala Ser Ser Arg Thr Ile Ala Thr Pro Val  
980 985 990

Gly Ala Leu Asn Pro Arg Trp Trp Ser Ala Ser Asp Asp Val Thr Val  
995 1000 1005

Tyr Pro Leu Pro Asp Gly Ala Thr Ser Leu Thr Pro Cys Thr Cys Gln  
1010 1015 1020

Ala Glu Ser Cys Trp Val Ile Arg Ser Asp Gly Ala Leu Cys His Gly  
1025 1030 1035 1040

Leu Ser Lys Gly Asp Lys Val Glu Leu Asp Val Ala Met Glu Val Ser  
1045 1050 1055

Asp Phe Arg Gly Ser Ser Gly Ser Pro Val Leu Cys Asp Glu Gly His  
1060 1065 1070

Ala Val Gly Met Leu Val Ser Val Leu His Ser Gly Gly Arg Val Thr  
1075 1080 1085

Ala Ala Arg Phe Thr Arg Pro Trp Thr Gln Val Pro Thr Asp Ala Lys  
1090 1095 1100

Thr Thr Thr Glu Pro Pro Pro Val Pro Ala Lys Gly Val Phe Lys Glu  
1105 1110 1115 1120

141

Ala Pro Leu Phe Met Pro Thr Gly Ala Gly Lys Ser Thr Arg Val Pro  
1125 1130 1135

Leu Glu Tyr Asp Asn Met Gly His Lys Val Leu Ile Leu Asn Pro Ser  
1140 1145 1150

Val Ala Thr Val Arg Ala Met Gly Pro Tyr Met Glu Arg Leu Ala Gly  
1155 1160 1165

Lys His Pro Ser Ile Tyr Cys Gly His Asp Thr Thr Ala Phe Thr Arg  
1170 1175 1180

Ile Thr Asp Ser Pro Leu Thr Tyr Ser Thr Tyr Gly Arg Phe Leu Ala  
1185 1190 1195 1200

Asn Pro Arg Gln Met Leu Arg Gly Val Ser Val Val Ile Cys Asp Glu  
1205 1210 1215

Cys His Ser His Asp Ser Thr Val Leu Leu Gly Ile Gly Arg Val Arg  
1220 1225 1230

Glu Leu Ala Arg Gly Cys Gly Val Gln Leu Val Leu Tyr Ala Thr Ala  
1235 1240 1245

Thr Pro Pro Gly Ser Pro Met Thr Gln His Pro Ser Ile Ile Glu Thr  
1250 1255 1260

Lys Leu Asp Val Gly Glu Ile Pro Phe Tyr Gly His Gly Ile Pro Leu  
1265 1270 1275 1280

Glu Arg Met Arg Thr Gly Arg His Leu Val Phe Cys His Ser Lys Ala  
1285 1290 1295

Glu Cys Glu Arg Leu Ala Gly Gln Phe Ser Ala Arg Gly Val Asn Ala  
1300 1305 1310

Ile Ala Tyr Tyr Arg Gly Lys Asp Ser Ser Ile Ile Lys Asp Gly Asp  
1315 1320 1325

Leu Val Val Cys Ala Thr Asp Ala Leu Ser Thr Gly Tyr Thr Gly Asn  
1330 1335 1340

Phe Asp Ser Val Thr Asp Cys Gly Leu Val Val Glu Glu Val Val Glu  
1345 1350 1355 1360

142

Val Thr Leu Asp Pro Thr Ile Thr Ile Ser Leu Arg Thr Val Pro Ala  
1365 1370 1375

Ser Ala Glu Leu Ser Met Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg  
1380 1385 1390

Ser Gly Arg Tyr Tyr Ala Gly Val Gly Lys Ala Pro Ala Gly Val  
1395 1400 1405

Val Arg Ser Gly Pro Val Trp Ser Ala Val Glu Ala Gly Val Thr Trp  
1410 1415 1420

Tyr Gly Met Glu Pro Asp Leu Thr Ala Asn Leu Leu Arg Leu Tyr Asp  
1425 1430 1435 1440

Asp Cys Pro Tyr Thr Ala Ala Val Ala Asp Ile Gly Glu Ala Ala  
1445 1450 1455

Val Phe Phe Ser Gly Leu Ala Pro Leu Arg Met His Pro Asp Val Ser  
1460 1465 1470

Trp Ala Lys Val Arg Gly Val Asn Trp Pro Leu Leu Val Gly Val Gln  
1475 1480 1485

Arg Thr Met Cys Arg Glu Thr Leu Ser Pro Gly Pro Ser Asp Asp Pro  
1490 1495 1500

Gln Trp Ala Gly Leu Lys Gly Pro Asn Pro Val Pro Leu Leu Leu Arg  
1505 1510 1515 1520

Trp Gly Asn Asp Leu Pro Ser Lys Val Ala Gly His His Ile Val Asp  
1525 1530 1535

Asp Leu Val Arg Arg Leu Gly Val Ala Glu Gly Tyr Val Arg Cys Asp  
1540 1545 1550

Ala Gly Pro Ile Leu Met Ile Gly Leu Ala Ile Ala Gly Gly Met Ile  
1555 1560 1565

Tyr Ala Ser Tyr Thr Gly Ser Leu Val Val Val Thr Asp Trp Asp Val  
1570 1575 1580

Lys Gly Gly Gly Ala Pro Leu Tyr Arg His Gly Asp Gln Ala Thr Pro  
1585 1590 1595 1600

143

Gln Pro Val Val Gln Val Pro Pro Val Asp His Arg Pro Gly Gly Glu  
1605 1610 1615

Ser Ala Pro Ser Asp Ala Lys Thr Val Thr Asp Ala Val Ala Ala Ile  
1620 1625 1630

Gln Val Asp Cys Asp Trp Thr Ile Met Thr Leu Ser Ile Gly Glu Val  
1635 1640 1645

Leu Ser Leu Ala Gln Ala Lys Thr Ala Glu Ala Tyr Thr Ala Thr Ala  
1650 1655 1660

Lys Trp Leu Ala Gly Cys Tyr Thr Gly Thr Arg Ala Val Pro Thr Val  
1665 1670 1675 1680

Ser Ile Val Asp Lys Leu Phe Ala Gly Gly Trp Ala Ala Val Val Gly  
1685 1690 1695

His Cys His Ser Val Ile Ala Ala Ala Val Ala Ala Tyr Gly Ala Ser  
1700 1705 1710

Arg Ser Pro Pro Leu Ala Ala Ala Ser Tyr Leu Met Gly Leu Gly  
1715 1720 1725

Val Gly Gly Asn Ala Gln Thr Arg Leu Ala Ser Ala Leu Leu Leu Gly  
1730 1735 1740

Ala Ala Gly Thr Ala Leu Gly Thr Pro Val Val Gly Leu Thr Met Ala  
1745 1750 1755 1760

Gly Ala Phe Met Gly Gly Ala Ser Val Ser Pro Ser Leu Val Thr Ile  
1765 1770 1775

Leu Leu Gly Ala Val Gly Gly Trp Glu Gly Val Val Asn Ala Ala Ser  
1780 1785 1790

Leu Val Phe Asp Phe Met Ala Gly Lys Leu Ser Ser Glu Asp Leu Trp  
1795 1800 1805

Tyr Ala Ile Pro Val Leu Thr Ser Pro Gly Ala Gly Leu Ala Gly Ile  
1810 1815 1820

Ala Leu Gly Leu Val Leu Tyr Ser Ala Asn Asn Ser Gly Thr Thr Thr  
1825 1830 1835 1840

144

Trp Leu Asn Arg Leu Leu Thr Thr Leu Pro Arg Ser Ser Cys Ile Pro  
1845 1850 1855

Asp Ser Tyr Phe Gln Gln Val Asp Tyr Cys Asp Lys Val Ser Ala Val  
1860 1865 1870

Leu Arg Arg Leu Ser Leu Thr Arg Thr Val Val Ala Leu Val Asn Arg  
1875 1880 1885

Glu Pro Lys Val Asp Glu Val Gln Val Gly Tyr Val Trp Asp Leu Trp  
1890 1895 1900

Glu Trp Ile Met Arg Gln Val Arg Val Val Met Ala Arg Leu Arg Ala  
1905 1910 1915 1920

Leu Cys Pro Val Val Ser Leu Pro Leu Trp His Cys Gly Glu Gly Trp  
1925 1930 1935

Ser Gly Glu Trp Leu Leu Asp Gly His Val Glu Ser Arg Cys Leu Cys  
1940 1945 1950

Gly Cys Val Ile Thr Gly Asp Val Leu Asn Gly Gln Leu Lys Glu Pro  
1955 1960 1965

Val Tyr Ser Thr Lys Leu Cys Arg His Tyr Trp Met Gly Thr Val Pro  
1970 1975 1980

Val Asn Met Leu Gly Tyr Gly Glu Thr Ser Pro Leu Leu Ala Ser Asp  
1985 1990 1995 2000

Thr Pro Lys Val Val Pro Phe Gly Thr Ser Gly Trp Ala Glu Val Val  
2005 2010 2015

Val Thr Thr Thr His Val Val Ile Arg Arg Thr Ser Ala Tyr Lys Leu  
2020 2025 2030

Leu Arg Gln Gln Ile Leu Ser Ala Ala Val Ala Glu Pro Tyr Tyr Val  
2035 2040 2045

Asp Gly Ile Pro Val Ser Trp Asp Ala Asp Ala Arg Ala Pro Ala Met  
2050 2055 2060

Val Tyr Gly Pro Gly Gln Ser Val Thr Ile Asp Gly Glu Arg Tyr Thr  
2065 2070 2075 2080

145

Leu Pro His Gln Leu Arg Leu Arg Asn Val Ala Pro Ser Glu Val Ser  
2085 2090 2095

Ser Glu Val Ser Ile Asp Ile Gly Thr Glu Thr Glu Asp Ser Glu Leu  
2100 2105 2110

Thr Glu Ala Asp Leu Pro Pro Ala Ala Ala Ala Leu Gln Ala Ile Glu  
2115 2120 2125

Asn Ala Ala Arg Ile Leu Glu Pro His Ile Asp Val Ile Met Glu Asp  
2130 2135 2140

Cys Ser Thr Pro Ser Leu Cys Gly Ser Ser Arg Glu Met Pro Val Trp  
2145 2150 2155 2160

Gly Glu Asp Ile Pro Arg Thr Pro Ser Pro Ala Leu Ile Ser Val Thr  
2165 2170 2175

Glu Ser Ser Ser Asp Glu Lys Thr Pro Ser Val Ser Ser Gln Glu  
2180 2185 2190

Asp Thr Pro Ser Ser Asp Ser Phe Glu Val Ile Gln Glu Ser Glu Thr  
2195 2200 2205

Ala Glu Gly Glu Glu Ser Val Phe Asn Val Ala Leu Ser Val Leu Lys  
2210 2215 2220

Ala Leu Phe Pro Gln Ser Asp Ala Thr Arg Lys Leu Thr Val Lys Met  
2225 2230 2235 2240

Ser Cys Cys Val Glu Lys Ser Val Thr Arg Phe Phe Ser Leu Gly Leu  
2245 2250 2255

Thr Val Ala Asp Val Ala Ser Leu Cys Glu Met Glu Ile Gln Asn His  
2260 2265 2270

Thr Ala Tyr Cys Asp Gln Val Arg Thr Pro Leu Glu Leu Gln Val Gly  
2275 2280 2285

Cys Leu Val Gly Asn Glu Leu Thr Phe Glu Cys Asp Lys Cys Glu Ala  
2290 2295 2300

Arg Gln Glu Thr Leu Ala Ser Phe Ser Tyr Ile Trp Ser Gly Val Pro  
2305 2310 2315 2320

146

Leu Thr Arg Ala Thr Pro Ala Lys Pro Pro Val Val Arg Pro Val Gly  
2325 2330 2335

Ser Leu Leu Val Ala Asp Thr Thr Lys Val Tyr Val Thr Asn Pro Asp  
2340 2345 2350

Asn Val Gly Arg Arg Val Asp Lys Val Thr Phe Trp Arg Ala Pro Arg  
2355 2360 2365

Val His Asp Lys Tyr Leu Val Asp Ser Ile Glu Arg Ala Lys Arg Ala  
2370 2375 2380

Ala Gln Ala Cys Leu Ser Met Gly Tyr Thr Tyr Glu Glu Ala Ile Arg  
2385 2390 2395 2400

Thr Val Arg Pro His Ala Ala Met Gly Trp Gly Ser Lys Val Ser Val  
2405 2410 2415

Lys Asp Leu Ala Thr Pro Ala Gly Lys Met Ala Val His Asp Arg Leu  
2420 2425 2430

Gln Glu Ile Leu Glu Gly Thr Pro Val Pro Phe Thr Leu Thr Val Lys  
2435 2440 2445

Lys Glu Val Phe Phe Lys Asp Arg Lys Glu Glu Lys Ala Pro Arg Leu  
2450 2455 2460

Ile Val Phe Pro Pro Leu Asp Phe Arg Ile Ala Glu Lys Leu Ile Leu  
2465 2470 2475 2480

Gly Asp Pro Gly Arg Val Ala Lys Ala Val Leu Gly Gly Ala Tyr Ala  
2485 2490 2495

Phe Gln Tyr Thr Pro Asn Gln Arg Val Lys Glu Met Leu Lys Leu Trp  
2500 2505 2510

Glu Ser Lys Lys Thr Pro Cys Ala Ile Cys Val Asp Ala Thr Cys Phe  
2515 2520 2525

Asp Ser Ser Ile Thr Glu Glu Asp Val Ala Leu Glu Thr Glu Leu Tyr  
2530 2535 2540

Ala Leu Ala Ser Asp His Pro Glu Trp Val Arg Ala Leu Gly Lys Tyr  
2545 2550 2555 2560

147

Tyr Ala Ser Gly Thr Met Val Thr Pro Glu Gly Val Pro Val Gly Glu  
2565 2570 2575

Arg Tyr Cys Arg Ser Ser Gly Val Leu Thr Thr Ser Ala Ser Asn Cys  
2580 2585 2590

Leu Thr Cys Tyr Ile Lys Val Lys Ala Ala Cys Glu Arg Val Gly Leu  
2595 2600 2605

Lys Asn Val Ser Leu Leu Ile Ala Gly Asp Asp Cys Leu Ile Ile Cys  
2610 2615 2620

Glu Arg Pro Val Cys Asp Pro Ser Asp Ala Leu Gly Arg Ala Leu Ala  
2625 2630 2635 2640

Ser Tyr Gly Tyr Ala Cys Glu Pro Ser Tyr His Ala Ser Leu Asp Thr  
2645 2650 2655

Ala Pro Phe Cys Ser Thr Trp Leu Ala Glu Cys Asn Ala Asp Gly Lys  
2660 2665 2670

Arg His Phe Phe Leu Thr Thr Asp Phe Arg Arg Pro Leu Ala Arg Met  
2675 2680 2685

Ser Ser Glu Tyr Ser Asp Pro Met Ala Ser Ala Ile Gly Tyr Ile Leu  
2690 2695 2700

Leu Tyr Pro Trp His Pro Ile Thr Arg Trp Val Ile Ile Pro His Val  
2705 2710 2715 2720

Leu Thr Cys Ala Phe Arg Gly Gly Thr Pro Ser Asp Pro Val Trp  
2725 2730 2735

Cys Gln Val His Gly Asn Tyr Tyr Lys Phe Pro Leu Asp Lys Leu Pro  
2740 2745 2750

Asn Ile Ile Val Ala Leu His Gly Pro Ala Ala Leu Arg Val Thr Ala  
2755 2760 2765

Asp Thr Thr Lys Thr Lys Met Glu Ala Gly Lys Val Leu Ser Asp Leu  
2770 2775 2780

Lys Leu Pro Gly Leu Ala Val His Arg Lys Lys Ala Gly Ala Leu Arg  
2785 2790 2795 2800

148

Thr Arg Met Leu Arg Ser Arg Gly Trp Ala Glu Leu Ala Arg Gly Leu  
2805 2810 2815

Leu Trp His Pro Gly Leu Arg Leu Pro Pro Pro Glu Ile Ala Gly Ile  
2820 2825 2830

Pro Gly Gly Phe Pro Leu Ser Pro Pro Tyr Met Gly Val Val His Gln  
2835 2840 2845

Leu Asp Phe Thr Ser Gln Arg Ser Arg Trp Arg Trp Leu Gly Phe Leu  
2850 2855 2860

Ala Leu Leu Ile Val Ala Leu Phe Gly  
2865 2870

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: PROBE 470-20-1-152F

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TCGGTTACTG AGAGCAGCTC AGATGAG

27

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid

149

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: JML-A, PRIMER

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

AGGAATTCAG CGGCCGCGAG

20

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: JML-B, PRIMER

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTCGCGGCCG CTGAATTCCCT TT

22

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 203 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

150

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: 470-20-1 CLONE, WITHOUT SISPA  
LINKERS

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 2..203

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

G GCT GTC TCG GAC TCT TGG ATG ACC TCG AAT GAG TCA GAG GAC GGG	46
Ala Val Ser Asp Ser Trp Met Thr Ser Asn Glu Ser Glu Asp Gly	
1                   5                   10                   15	
GTA TCC TCC TGC GAG GAG GAC ACC GGC GGG GTC TTC TCA TCT GAG CTG	94
Val Ser Ser Cys Glu Glu Asp Thr Gly Gly Val Phe Ser Ser Glu Leu	
20                25                30	
CTC TCA GTA ACC GAG ATA AGT GCT GCC GAT GGA GTA CGG GGG ATG TCT	142
Leu Ser Val Thr Glu Ile Ser Ala Gly Asp Gly Val Arg Gly Met Ser	
35                40                45	
TCT CCC CAT ACA GGC ATC TCT CGG CTA CTA CCA CAA AGA GAG GGT GTA	190
Ser Pro His Thr Gly Ile Ser Arg Leu Leu Pro Gln Arg Glu Gly Val	
50                55                60	
CTG CAG TCC TCC A	203
Leu Gln Ser Ser	
65	

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 67 amino acids
- (B) TYPE: amino acid
- (D) TOP LOGY: linear

(ii) MOLECULE TYPE: protein

151

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Ala Val Ser Asp Ser Trp Met Thr Ser Asn Glu Ser Glu Asp Gly Val  
1 5 10 15

Ser Ser Cys Glu Glu Asp Thr Gly Gly Val Phe Ser Ser Glu Leu Leu  
20 25 30

Ser Val Thr Glu Ile Ser Ala Gly Asp Gly Val Arg Gly Met Ser Ser  
35 40 45

Pro His Thr Gly Ile Ser Arg Leu Leu Pro Gln Arg Glu Gly Val Leu  
50 55 60

Gln Ser Ser  
65

## (2) INFORMATION FOR SEQ ID NO:21:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: 470-20-1-152R

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CTCATCTGAG CTGCTCTCAG TAACCGA

27

## (2) INFORMATION FOR SEQ ID NO:22:

## (i) SEQUENCE CHARACTERISTICS:

152

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: OLIGONUCLEOTIDE B

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CTGTCTCGGA CTCTTGGATG ACCT

24

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 24 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: COGNATE OLIGONUCLEOTIDE 211R'

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

ATACCCCGTC CTCTGACTCA TTCTG

24

(2) INFORMATION FOR SEQ ID NO:24:

153

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: COGNATE OLIGONUCLEOTIDE B'

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

AGGTCATCCA AGAGTCCGAG ACAG

24

## (2) INFORMATION FOR SEQ ID NO:25:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: LAMBDA GT 11 FORWARD PRIMER, 20mer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CACATGGCTG AATATCGACG

20

154

## (2) INFORMATION FOR SEQ ID NO:26:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: PROBE 470-201-1-142R

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TCGGTTACTG AGAGCAGCTC AGATGAG

27

## (2) INFORMATION FOR SEQ ID NO:27:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: PROBE 470-20-1-152F

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TCGGTTACTG AGAGCAGCTC AGATGAG

27

## (2) INFORMATION FOR SEQ ID NO:28:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 570 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Clone 470EXP1

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..570

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCT GTA TGG TTC TCG ATT TCC ATC TCA CAC AGG CTA GCA ACA TCA GCC Ala Val Trp Phe Trp Ile Ser Ile Ser His Arg Leu Ala Thr Ser Ala	48
1 5 10 15	
ACC GTC AAC CCC AAT GAG AAA AAG CGC GTG ACG CTC TTT TCA ACG CAG Thr Val Asn Pro Asn Glu Lys Lys Arg Val Thr Leu Phe Ser Thr Gln	96
20 25 30	
CAC GAC ATC TTG ACG GTA AGC TTC CTG GTC GCG TCG CTC TGT GGA AAT His Asp Ile Leu Thr Val Ser Phe Leu Val Ala Ser Leu Cys Gly Asn	144
35 40 45	
AAG GCT TTT AAT ACG GAA AGA GCC ACG TTG AAG ACA CTT TCC TCC CCT Lys Ala Phe Asn Thr Glu Arg Ala Thr Leu Lys Thr Leu Ser Ser Pro	192
50 55 60	
TCG GCT GTC TCG GAC TCT TGG ATG ACC TCG AAT GAG TCA GAG GAC GGG Ser Ala Val Ser Asp Ser Trp Met Thr Ser Asn Glu Ser Glu Asp Gly	240
65 70 75 80	

156

GTA TCC TCC TGC GAG GAG GAC ACC GAC GGG GTC TTC TCA TCT GAG CTG	288
Val Ser Ser Cys Glu Glu Asp Thr Asp	ly Val Phe Ser Ser Glu Leu
85	90
	95
CTC TCA GTA ACC GAG ATA AGT GCT GGC GAT GGA GTA CGG GGG ATG TCT	336
Leu Ser Val Thr Glu Ile Ser Ala Gly Asp Gly Val Arg Gly Met Ser	
100	105
	110
TCT CCC CAT ACA GGC ATC TCT CGG CTA CTA CCA CAA AGA GAG GGT GTA	384
Ser Pro His Thr Gly Ile Ser Arg Leu Leu Pro Gln Arg Glu Gly Val	
115	120
	125
CTG CAG TCC TCC ATG ATG ACA TCA ATG TGC GGT TCA AGA ATC CTC GCA	432
Leu Gln Ser Ser Met Met Thr Ser Met Cys Gly Ser Arg Ile Leu Ala	
130	135
	140
GCA TTC TCG ATC GCT TGG AGA GCA GCA GCC GCC GGC AGA TCG GCC	480
Ala Phe Ser Ile Ala Trp Arg Ala Ala Ala Gly Gly Arg Ser Ala	
145	150
	155
	160
TCA GTC AGT TCT GAG TCT TCA GTC TCC GTC CCA ATG TCA ATG GAC ACC	528
Ser Val Ser Ser Glu Ser Ser Val Ser Val Pro Met Ser Met Asp Thr	
165	170
	175
TCG GAT GAA ACC TCA GAG GGT GCC ACA TTC CTG AGC CTC AGT	570
Ser Asp Glu Thr Ser Glu Gly Ala Thr Phe Leu Ser Leu Ser	
180	185
	190

## (2) INFORMATION FOR SEQ ID NO:29:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 190 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Ala Val Trp Phe Trp Ile Ser Ile Ser His Arg Leu Ala Thr Ser Ala  
 1                   5                   10                   15

Thr Val Asn Pro Asn Glu Lys Lys Arg Val Thr Leu Phe Ser Thr Gln

157

20	25	30
<b>His Asp Ile Leu Thr Val Ser Phe Leu Val Ala Ser Leu Cys Gly Asn</b>		
35	40	45
<b>Lys Ala Phe Asn Thr Glu Arg Ala Thr Leu Lys Thr Leu Ser Ser Pro</b>		
50	55	60
<b>Ser Ala Val Ser Asp Ser Trp Met Thr Ser Asn Glu Ser Glu Asp Gly</b>		
65	70	75
<b>Val Ser Ser Cys Glu Glu Asp Thr Asp Gly Val Phe Ser Ser Glu Leu</b>		
85	90	95
<b>Leu Ser Val Thr Glu Ile Ser Ala Gly Asp Gly Val Arg Gly Met Ser</b>		
100	105	110
<b>Ser Pro His Thr Gly Ile Ser Arg Leu Leu Pro Gln Arg Glu Gly Val</b>		
115	120	125
<b>Leu Gln Ser Ser Met Met Thr Ser Met Cys Gly Ser Arg Ile Leu Ala</b>		
130	135	140
<b>Ala Phe Ser Ile Ala Trp Arg Ala Ala Ala Gly Gly Arg Ser Ala</b>		
145	150	155
<b>160</b>		
<b>Ser Val Ser Ser Glu Ser Ser Val Ser Val Pro Met Ser Met Asp Thr</b>		
165	170	175
<b>Ser Asp Glu Thr Ser Glu Gly Ala Thr Phe Leu Ser Leu Ser</b>		
180	185	190

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

158

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer GE-3F

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GCCGCCATGG TCTCATGGGA CGCGGACGCT CGTGCGCCCG CGATG

45

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer GE-3R

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GCGCGGATCC GATAAGTGCT GGCGATGGAG TACG

34

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

159

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer GE-9F

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GGCACCATGG TCACCCCGGA AG

22

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer GE-9R

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GCTCGGATCC GGAGCAGAAC GGGGCCGT

28

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 364 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

160

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: GE3-2

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 2..364

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

G GTC TCA TGG GAC GCG GAC GCT CGT GCG CCC GCG ATG GTC TAT GGC	46
Val Ser Trp Asp Ala Asp Ala Arg Ala Pro Ala Met Val Tyr Gly	
1 5 10 15	
CCT GGG CAA AGT GTT ACC ATT GAC GGG GAG CGC TAC ACC TTG CCT CAT	94
Pro Gly Gln Ser Val Thr Ile Asp Gly Glu Arg Tyr Thr Leu Pro His	
20 25 30	
CAA CTG AGG CTC AGG AAT GTG GCA CCC TCT GAG GTT TCA TCC GAG GTG	142
Gln Leu Arg Leu Arg Asn Val Ala Pro Ser Glu Val Ser Ser Glu Val	
35 40 45	
TCC ATT GAC ATT GGG ACG GAG ACT GAA GAC TCA GAA CTG ACT GAG GCC	190
Ser Ile Asp Ile Gly Thr Glu Thr Glu Asp Ser Glu Leu Thr Glu Ala	
50 55 60	
GAT CTG CCG CCG GCG GCT GCT CTC CAA GCG ATC GAG AAT GCT GCG	238
Asp Leu Pro Pro Ala Ala Ala Leu Gln Ala Ile Glu Asn Ala Ala	
65 70 75	
AGG ATT CTT GAA CCG CAC ATT GAT GTC ATC ATG GAG GAC TGC AGT ACA	286
Arg Ile Leu Glu Pro His Ile Asp Val Ile Met Glu Asp Cys Ser Thr	
80 85 90 95	
CCC TCT CTT TGT GGT AGT AGC CGA GAG ATG CCT GTA TGG GGA GAA GAC	334
Pro Ser Leu Cys Gly Ser Ser Arg Glu Met Pro Val Trp Gly Glu Asp	
100 105 110	

161

ATC CCC CGT ACT CCA TCG CCA GCA CTT ATC  
Ile Pro Arg Thr Pro Ser Pro Ala Leu Ile  
115 120

364

## (2) INFORMATION FOR SEQ ID NO:35:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 121 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Val Ser Trp Asp Ala Asp Ala Arg Ala Pro Ala Met Val Tyr Gly Pro  
1 5 10 15

Gly Gln Ser Val Thr Ile Asp Gly Glu Arg Tyr Thr Leu Pro His Gln  
20 25 30

Leu Arg Leu Arg Asn Val Ala Pro Ser Glu Val Ser Ser Glu Val Ser  
35 40 45

Ile Asp Ile Gly Thr Glu Thr Glu Asp Ser Glu Leu Thr Glu Ala Asp  
50 55 60

Leu Pro Pro Ala Ala Ala Leu Gln Ala Ile Glu Asn Ala Ala Arg  
65 70 75 80

Ile Leu Glu Pro His Ile Asp Val Ile Met Glu Asp Cys Ser Thr Pro  
85 90 95

Ser Leu Cys Gly Ser Ser Arg Glu Met Pro Val Trp Gly Glu Asp Ile  
100 105 110

Pro Arg Thr Pro Ser Pro Ala Leu Ile  
115 120

## (2) INFORMATION FOR SEQ ID NO:36:

## (i) SEQUENCE CHARACTERISTICS:

162

- (A) LENGTH: 290 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Clone GE9-2

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 3..290

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CC ATG GTC ACC CCG GAA GGG GTG CCC GTT GGT GAG AGG TAT TGC AGA	47
Met Val Thr Pro Glu Gly Val Pro Val Gly Glu Arg Tyr Cys Arg	
1                   5                   10                   15	

TCC TCG GGT GTC CTA ACA ACT AGC GCG AGC AAC TGC TTG ACC TGC TAC	95
Ser Ser Gly Val Leu Thr Thr Ser Ala Ser Asn Cys Leu Thr Cys Tyr	
20                25                30	

ATC AAG GTG AAA GCC GCC TGT GAG AGG GTG GGG CTG AAA AAT GTC TCT	143
Ile Lys Val Lys Ala Ala Cys Glu Arg Val Gly Leu Lys Asn Val Ser	
35               40                45	

CTT CTC ATA GCC GGC GAT GAC TGC TTG ATC ATA TGT GAG CGG CCA GTG	191
Leu Leu Ile Ala Gly Asp Asp Cys Leu Ile Ile Cys Glu Arg Pro Val	
50               55                60	

TGC GAC CCA AGC GAC GCT TTG GGC AGA GCC CTA GCG AGC TAT GGG TAC	239
Cys Asp Pro Ser Asp Ala Leu Gly Arg Ala Leu Ala Ser Tyr Gly Tyr	
65               70                75	

GCG TGC GAG CCC TCA TAT TAT GCA TGC TCG GAC ACG GCC CCC TTC TGC	287
Ala Cys Glu Pro Ser Tyr Tyr Ala Cys Ser Asp Thr Ala Pro Phe Cys	
80               85                90                95	

163

TCC  
Ser

290

## (2) INFORMATION FOR SEQ ID NO:37:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 96 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Met Val Thr Pro Glu Gly Val Pro Val Gly Glu Arg Tyr Cys Arg Ser  
1 5 10 15

Ser Gly Val Leu Thr Thr Ser Ala Ser Asn Cys Leu Thr Cys Tyr Ile  
20 25 30

Lys Val Lys Ala Ala Cys Glu Arg Val Gly Leu Lys Asn Val Ser Leu  
35 40 45

Leu Ile Ala Gly Asp Asp Cys Leu Ile Ile Cys Glu Arg Pro Val Cys  
50 55 60

Asp Pro Ser Asp Ala Leu Gly Arg Ala Leu Ala Ser Tyr Gly Tyr Ala  
65 70 75 80

Cys Glu Pro Ser Tyr Tyr Ala Cys Ser Asp Thr Ala Pro Phe Cys Ser  
85 90 95

## (2) INFORMATION FOR SEQ ID NO:38:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA to mRNA

164

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: JML-A SISPA Primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

AGGAATTCAG CGGCCGCGAG

20

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: JML-B SISPA Primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CTCGCGGCCG CTGAATTCCCT TT

22

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

165

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: 470ep-f1 Primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GCGAATTCCGC CATGGCGGGG AGACTTTCAT CA

32

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: 470ep-R1 Primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GCGAATTCCGG ATCCAGGGCC ATAGACCATC GCGGG

35

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both

166

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: 470ep-f2 Primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GCGAATTCCG TGCGCCGCC ATGGTC

26

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: 470ep-R3 Primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GCGAATTCCG ATCCCAAGGT TTCTTGCTA GC

32

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs

167

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: 470ep-f4 Primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GCGAATTCAA GTGTGAGGCT AGGCAA

26

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: 470ep-R4 Primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GCGAATTCCGG ATCCCCACAC AGATGGCGCA AGGGG

35

(2) INFORMATION FOR SEQ ID NO:46:

168

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: KL-1 SISPA Primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

GCAGGATCCG AATTCCGCATC TAGAGAT

27

## (2) INFORMATION FOR SEQ ID NO:47:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: KL-2 SISPA Primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

ATCTCTAGAT GCGAATTCGG ATCCTGCGA

29

## (2) INFORMATION FOR SEQ ID NO:48:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 186 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Clone Y5-10

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..186

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CGT GCG CCC GCC ATG GTC TAT GGC CCT GGG CAA AGT GTT GCC ATT GAC	48
Arg Ala Pro Ala Met Val Tyr Gly Pro Gly Gln Ser Val Ala Ile Asp	
1                   5                   10                   15	
GGG GAG CGC TAC ACC TTG CCT CAT CAA CTG AGG CTC AGG AAT GTG GCA	96
Gly Glu Arg Tyr Thr Leu Pro His Gln Leu Arg Leu Arg Asn Val Ala	
20               25               30	
CCC TCT GAG GTT TCA TCC GAG GTG TCC ATT GAC ATT GGG ACC GAG GCT	144
Pro Ser Glu Val Ser Ser Glu Val Ser Ile Asp Ile Gly Thr Glu Ala	
35               40               45	
GAA AAC TCA GAA CTG ACT GAG GCC GAT CTG CCG CCG GCG GCT	186
Glu Asn Ser Glu Leu Thr Glu Ala Asp Leu Pro Pro Ala Ala	
50               55               60	

(2) INFORMATION FOR SEQ ID NO:49:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 62 amino acids
- (B) TYPE: amino acid

170

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

Arg Ala Pro Ala Met Val Tyr Gly Pro Gly Gln Ser Val Ala Ile Asp  
1 5 10 15

Gly Glu Arg Tyr Thr Leu Pro His Gln Leu Arg Leu Arg Asn Val Ala  
20 25 30

Pro Ser Glu Val Ser Ser Glu Val Ser Ile Asp Ile Gly Thr Glu Ala  
35 40 45

Glu Asn Ser Glu Leu Thr Glu Ala Asp Leu Pro Pro Ala Ala  
50 55 60

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 282 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Clone Y5-12

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..282

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

CGT GCG CCC GCC ATG GTC TAT GGC CCT GGG CAA ACT GTT ACC ATT GAC

48

171

Arg Ala Pro Ala Met Val Tyr Gly Pro Gly Gln Ser Val Thr Ile Asp			
1	5	10	15
GGG GAG CGC TAC ACC TTG CCT CAT CAA CTG AGG CTC AGG AAT GTG GCA			96
Gly Glu Arg Tyr Thr Leu Pro His Gln Leu Arg Leu Arg Asn Val Ala			
20	25	30	
CCC TCT GAG GTT TCA TCC GAG GTG TCC ATT GAC ATT GGG ACG GAG ACT			144
Pro Ser Glu Val Ser Ser Glu Val Ser Ile Asp Ile Gly Thr Glu Thr			
35	40	45	
GAA GAC TCA GAA CTG ACT GAG GCC GAT CTG CCG CCG GCG GCT GCT GCT			192
Glu Asp Ser Glu Leu Thr Glu Ala Asp Leu Pro Pro Ala Ala Ala Ala			
50	55	60	
CTC CAA GCG ATC GAG AAT GCT GCG AGG ATT CTT GAA CCG CAC ATT GAT			240
Leu Gln Ala Ile Glu Asn Ala Ala Arg Ile Leu Glu Pro His Ile Asp			
65	70	75	80
GTC ATC ATG GAG GAC TGC AGT ACA CCC TCT CTT TGT GGT AGT			282
Val Ile Met Glu Asp Cys Ser Thr Pro Ser Leu Cys Gly Ser			
85	90		

## (2) INFORMATION FOR SEQ ID NO:51:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 94 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

Arg Ala Pro Ala Met Val Tyr Gly Pro Gly Gln Ser Val Thr Ile Asp			
1	5	10	15
Gly Glu Arg Tyr Thr Leu Pro His Gln Leu Arg Leu Arg Asn Val Ala			
20	25	30	
Pro Ser Glu Val Ser Ser Glu Val Ser Ile Asp Ile Gly Thr Glu Thr			
35	40	45	

172

Glu Asp Ser Glu Leu Thr Glu Ala Asp Leu Pro Pro Ala Ala Ala Ala  
 50                    55                    60

Leu Gln Ala Ile Glu Asn Ala Ala Arg Ile Leu Glu Pro His Ile Asp  
 65                    70                    75                    80

Val Ile Met Glu Asp Cys Ser Thr Pro Ser Leu Cys Gly Ser  
 85                    90

## (2) INFORMATION FOR SEQ ID NO:52:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 279 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Clone Y5-26

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..279

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

CGT GCG CCC GCC ATG GTC TAT GGC CCT GGG CAA AGT GTT TCC ATT GAC            48  
 Arg Ala Pro Ala Met Val Tyr Gly Pro Gly Gln Ser Val Ser Ile Asp  
 1                    5                    10                    15

GGG GAG CGC TAC ACC TTG CCT CAT CAA CTG AGG CTC AGG AAT GTG GCA            96  
 Gly Glu Arg Tyr Thr Leu Pro His Gln Leu Arg Leu Arg Asn Val Ala  
 20                    25                    30

CCC TCT GAG GTT TCA TCC GAG GTG TCC ATT GAC ATT GGG ACG GAG ACT            144  
 Pro Ser Glu Val Ser Ser Glu Val Ser Ile Asp Ile Gly Thr Glu Thr

173

35

40

45

GAA GAC TCA GAA CTG ACT GAG GCC GAC CTG CCG CCG GCG GCT GCT GCT      192  
 Glu Asp Ser Glu Leu Thr Glu Ala Asp Leu Pro Pro Ala Ala Ala Ala

50

55

60

CTC CAA GCG ATC GAG AAT GCT GCG AGG ATT CTT GAA CCG CAC ATC GAT      240  
 Leu Gln Ala Ile Glu Asn Ala Ala Arg Ile Leu Glu Pro His Ile Asp  
 65                    70                    75                    80

GTC ATC ATG GAG GAC TGC AGT ACA CCC TCT CTT TGT GGT      279  
 Val Ile Met Glu Asp Cys Ser Thr Pro Ser Leu Cys Gly  
 85                    90

## (2) INFORMATION FOR SEQ ID NO:53:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 93 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

Arg Ala Pro Ala Met Val Tyr Gly Pro Gly Gln Ser Val Ser Ile Asp

1

5

10

15

Gly Glu Arg Tyr Thr Leu Pro His Gln Leu Arg Leu Arg Asn Val Ala

20

25

30

Pro Ser Glu Val Ser Ser Glu Val Ser Ile Asp Ile Gly Thr Glu Thr

35

40

45

Glu Asp Ser Glu Leu Thr Glu Ala Asp Leu Pro Pro Ala Ala Ala

50

55

60

Leu Gln Ala Ile Glu Asn Ala Ala Arg Ile Leu Glu Pro His Ile Asp

65

70

75

80

Val Ile Met Glu Asp Cys Ser Thr Pro Ser Leu Cys Gly

85

90

174

## (2) INFORMATION FOR SEQ ID NO:54:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 108 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Clone Y5-5

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..108

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

GCC TAT TGT GAC AAG GTG CGC ACT CCG CTT GAA TTG CAG GTT GGG TGC	48		
Ala Tyr Cys Asp Lys Val Arg Thr Pro Leu Glu Leu Gln Val Gly Cys			
1	5	10	15

TTG GTG GGC AAT GAA CTT ACC TTT GAA TGT GAC AAG TGT GAG GCT AGG	96		
Leu Val Gly Asn Glu Leu Thr Phe Glu Cys Asp Lys Cys Glu Ala Arg			
20	25	30	

CAA GAA ACC TTG	108
Gln Glu Thr Leu	
35	

## (2) INFORMATION FOR SEQ ID NO:55:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

175

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Ala Tyr Cys Asp Lys Val Arg Thr Pro Leu Glu Leu Gln Val Gly Cys  
 1               5                           10                           15

Leu Val Gly Asn Glu Leu Thr Phe Glu Cys Asp Lys Cys Glu Ala Arg  
 20   25                           30

Gln Glu Thr Leu  
 35

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 132 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Clone Y5-3

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..132

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

GAG ATG GAA ATC CAG AAC CAT ACA GCC TAT TGT GAC AAG GTG CGC ACT  
 Glu Met Glu Ile Gln Asn His Thr Ala Tyr Cys Asp Lys Val Arg Thr  
 1               5                           10                           15

48

CCG CTT GAA TTG CAG GTT GGG TGC TTG GTG GGC AAT GAA CTT ACC TTT  
 Pro Leu Glu Leu Gln Val Gly Cys Leu Val Gly Asn Glu Leu Thr Phe

96

176

20

25

30

GAA TGT GAC AAG TGT GAG GCT AGG CAA GAA ACC TTG  
Glu Cys Asp Lys Cys Glu Ala Arg Gln Glu Thr Leu

35

40

132

## (2) INFORMATION FOR SEQ ID NO:57:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(iii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

Glu Met Glu Ile Gln Asn His Thr Ala Tyr Cys Asp Lys Val Arg Thr  
1 5 10 15

Pro Leu Glu Leu Gln Val Gly Cys Leu Val Gly Asn Glu Leu Thr Phe  
20 25 30

Glu Cys Asp Lys Cys Glu Ala Arg Gln Glu Thr Leu  
35 40

## (2) INFORMATION FOR SEQ ID NO:58:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 258 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Clone Y5-27

177

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..258

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

AAA GCC TTA TTT CCA CAG AGC GAC GCG ACC AGG AAG CTT ACC GTC AAG	48
Lys Ala Leu Phe Pro Gln Ser Asp Ala Thr Arg Lys Leu Thr Val Lys	
1                   5                   10                   15	
ATG TCA TGC TGC GTT GAA AAG AGC GTC ACG CGC TTT TTC TCA TTG GGG	96
Met Ser Cys Cys Val Glu Lys Ser Val Thr Arg Phe Phe Ser Leu Gly	
20                   25                   30	
TTG ACG GTG GCT GAT GTT GCT AGC CTG TGT GAG ATG GAA ATC CAG AAC	144
Leu Thr Val Ala Asp Val Ala Ser Leu Cys Glu Met Glu Ile Gln Asn	
35                   40                   45	
CAT ATA GCC TAT TGT GAC AAG GTG CGC ACT CCG CTT GAA TTG CAG GTT	192
His Ile Ala Tyr Cys Asp Lys Val Arg Thr Pro Leu Glu Leu Gln Val	
50                   55                   60	
GGG TGC TTG GTG GGC AAT GAA CTC ACC TTT GAA TGT GAC AAG TGT GAG	240
Gly Cys Leu Val Gly Asn Glu Leu Thr Phe Glu Cys Asp Lys Cys Glu	
65                   70                   75                   80	
GCT AGG CAA GAA ACC TTG	258
Ala Arg Gln Glu Thr Leu	
85	

## (2) INFORMATION FOR SEQ ID NO:59:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 86 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

Lys Ala Leu Phe Pro Gln Ser Asp Ala Thr Arg Lys Leu Thr Val Lys

178

1                   5                   10                   15

Met Ser Cys Cys Val Glu Lys Ser Val Thr Arg Phe Phe Ser Leu Gly  
 20                   25                   30

Leu Thr Val Ala Asp Val Ala Ser Leu Cys Glu Met Glu Ile Gln Asn  
 35                   40                   45

His Ile Ala Tyr Cys Asp Lys Val Arg Thr Pro Leu Glu Leu Gln Val  
 50                   55                   60

Gly Cys Leu Val Gly Asn Glu Leu Thr Phe Glu Cys Asp Lys Cys Glu  
 65                   70                   75                   80

Ala Arg Gln Glu Thr Leu  
 85

**(2) INFORMATION FOR SEQ ID NO:60:**

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 108 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

**(ii) MOLECULE TYPE:** cDNA to mRNA

**(iii) HYPOTHETICAL:** NO

**(iv) ANTI-SENSE:** NO

**(v.i) ORIGINAL SOURCE:**

(C) INDIVIDUAL ISOLATE: Clone Y5-25

**(ix) FEATURE:**

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..108

**(xi) SEQUENCE DESCRIPTION:** SEQ ID NO:60:

ACC TAT TGT GAC AAG GTG CGC ACT CCG CTT GAA TTG CAG GTT GGG TGC       48  
 Thr Tyr Cys Asp Lys Val Arg Thr Pro Leu Glu Leu Gln Val Gly Cys

179

1 . . . . . 5 . . . . . 10 . . . . . 15

15

TTG GTG GGC AAT GAA CTT ACC TTT GAA TGT GAC AAG TGT GAG GCT AGG  
Leu Val Gly Asn Glu Leu Thr Phe Glu Cys Asp Lys Cys Glu Ala Arg

30

85

10

96

**CAA GAA ACC TTG**

108

35

(2) INFORMATION FOR SEO ID NO: 61:

### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

Thr Tyr Cys Asp Lys Val Arg Thr Pro Leu Glu Leu Gln Val Gly Cys  
 1                5                10                15

1

5

2

1

**Leu Val Gly Asn Glu Leu Thr Phe Glu Cys Asp Lys Cys Glu Ala Arg**

30

25

四

Gln Glu Thr Leu

35

(2) INFORMATION FOR SEQ ID NO:62:

#### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 108 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: both
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

180

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(c) INDIVIDUAL ISOLATE: Clone Y5-20

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 52..108

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

GCCGACACTA CTAAGGTGTA TGTTACCAAT CCAGACAATG TGGGACGAAG G GTG GGC	57
	Val Gly
	1

AAT GAA CTT ACC TTT GAA TGT GAC AAG TGT GAG GCT AGG CAA GAA ACC	105	
Asn Glu Leu Thr Phe Glu Cys Asp Lys Cys Glu Ala Arg Gln Glu Thr		
5	10	15

TTG	108
Leu	

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

Val	Gly	Asn	Glu	Leu	Thr	Phe	Glu	Cys	Asp	Lys	Cys	Glu	Ala	Arg	Gln
1															

1	5	10	15
---	---	----	----

Glu Thr Leu

(2) INFORMATION FOR SEQ ID NO:64:

181

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 168 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Clone Y5-16

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..168

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

TTG GGG TTG ACG GTG GCT GAT GTT GCT AGC CTG TGT GAG ATG GAA ATC	48
Leu Gly Leu Thr Val Ala Asp Val Ala Ser Leu Cys Glu Met Glu Ile	
1                   5                   10                   15	
CAG AAC CAT ACA GCC TAT TGT GAC AAG GTG CGC ACT CCG CTT GAA TTG	96
Gln Asn His Thr Ala Tyr Cys Asp Lys Val Arg Thr Pro Leu Glu Leu	
20                25                30	
CAG GTT GGG TGC TTG GTG GGC AAT GAA CTT ACC TTT GAA TGT GAC AAG	144
Gln Val Gly Cys Leu Val Gly Asn Glu Leu Thr Phe Glu Cys Asp Lys	
35                40                45	
TGT GAG GCT ACG CAA GAA ACC TTG	168
Cys Glu Ala Arg Gln Glu Thr Leu	
50                55	

(2) INFORMATION FOR SEQ ID NO:65:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 56 amino acids
- (B) TYPE: amino acid

182

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

Gln Asn His Thr Ala Tyr Cys Asp Lys Val Arg Thr Pro Leu Glu Leu  
20 25 30

Gln Val Gly Cys Leu Val Gly Asn Glu Leu Thr Phe Glu Cys Asp Lys  
35 40 45

Cys Glu Ala Arg Gln Glu Thr Leu  
50 55

(2) INFORMATION FOR SEQ ID NO:66:

**(i) SEQUENCE CHARACTERISTICS:**

(A) LENGTH: 313 base pairs

(B) TYPE: nucleic acid

**(C) STRANDEDNESS:** bot

## ORIGINAL SOURCE:

#### **6.1.2. 财政政策**

#### **FEATURES.**

(B) LOCATION: 3-312

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66.

ATC ACC GTC AAC CCC AAT GAG AAA AAG CGC GTG ACG CTC TTT TCA ACG  
Ile Thr Val Asn Pro Asn Glu Lys Lys Arg Val Thr Leu Phe Ser Thr

			183
1	5	10	15
CAG CAC GAC ATC TTG ACG GTA AGC TTC CTG GTC GCG TCG CTC TGT GGA			96
Gln His Asp Ile Leu Thr Val Ser Phe Leu Val Ala Ser Leu Cys Gly			
20	25	30	
AAT AAG GCT TTT AAT ACG GAA AGA GCC ACG TTG AAG ACA CTT TCC TCC			144
Asn Lys Ala Phe Asn Thr Glu Arg Ala Thr Leu Lys Thr Leu Ser Ser			
35	40	45	
CCT TCG GCT GTC TCG GAC TCT TGG ATG ACC TCG AAT GAG TCA GAG GAC			192
Pro Ser Ala Val Ser Asp Ser Trp Met Thr Ser Asn Glu Ser Glu Asp			
50	55	60	
GGG GTA TCC TCC TGC GAG GAG GAC ACC GAC GGG GTC TTC TCA TCT GAG			240
Gly Val Ser Ser Cys Glu Glu Asp Thr Asp Gly Val Phe Ser Ser Glu			
65	70	75	80
CTG CTC TCA GTA ACC GAG ATA AGT GCT GGC GAT GGA GTA CGG GGG ATG			288
Leu Leu Ser Val Thr Glu Ile Ser Ala Gly Asp Gly Val Arg Gly Met			
85	90	95	
TCT TCT CCC CAT ACA GGC ATC TCT C			313
Ser Ser Pro His Thr Gly Ile Ser			
100			

**(2) INFORMATION FOR SEQ ID NO:67:**

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 104 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

184

Asn Lys Ala Phe Asn Thr Glu Arg Ala Thr Leu Lys Thr Leu Ser Ser			
35	40	45	
Pro Ser Ala Val Ser Asp Ser Trp Met Thr Ser Asn Glu Ser Glu Asp			
50	55	60	
Gly Val Ser Ser Cys Glu Glu Asp Thr Asp Gly Val Phe Ser Ser Glu			
65	70	75	80
Leu Leu Ser Val Thr Glu Ile Ser Ala Gly Asp Gly Val Arg Gly Met			
85	90	95	
Ser Ser Pro His Thr Gly Ile Ser			
100			

## (2) INFORMATION FOR SEQ ID NO:68:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 89 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Clone Y5-52

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 28..87

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

ACTGAGAGCA GCTCAGATGA GAAGACC CCT TCG GCT GTC TCG GAC TCT TGG	
Pro Ser Ala Val Ser Asp Ser Trp	

51

1

5

185

ATG ACC TCG AAT GAG TCA GAG GAC GGG GTA TCC TCG CA  
Met Thr Ser Asn Glu Ser Glu Asp Gly Val Ser Ser  
10 15 20

89

## (2) INFORMATION FOR SEQ ID NO:69:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

Pro Ser Ala Val Ser Asp Ser Trp Met Thr Ser Asn Glu Ser Glu Asp  
1 5 10 15  
  
Gly Val Ser Ser  
20

## (2) INFORMATION FOR SEQ ID NO:70:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 214 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Clone Y5-53

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..100

186

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

AAT AAG GCT TTT AAT ACG GAA AGA GCC ACG TTG AAG ACA CTT TCC TCC  
 Asn Lys Ala Phe Asn Thr Glu Arg Ala Thr Leu Lys Thr Leu Ser Ser  
 1 5 10 15

CCT TCG GCT GTC TCG GAC TCT TGG ATG ACC TCG AAT GAG TCA GAG GAC 96  
 Pro Ser Ala Val Ser Asp Ser Trp Met Thr Ser Asn Glu Ser Glu Asp  
           20                     25                     30

GGG G ATCTCTAGAT GCGAATTCAA GTGTGAGGCT AGGCAAGAAA CCTTGGCCTC 150  
Gly

**CTTCTCTTAC ATTTGGCTCG GAGTGCCGCT GACTAGGGCC ACGCCCCCCTA AGCCCTCCCGT** 210

**GGTG** 314

(2) INFORMATION FOR SEO ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:-

- (A) LENGTH: 33 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

Asn Lys Ala Phe Asn Thr Glu Arg Ala Thr Leu Lys Thr Leu Ser Ser  
1 5 10 15

Pro Ser Ala Val Ser Asp Ser Trp Met Thr Ser Asn Glu Ser Glu Asp  
20 25 30

Gly

(2) INFORMATION FOR SEQ ID NO:72:

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 113 base pairs
  - (B) TYPE: nucleic acid

187

- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Clone Y5-55

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 52..113

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

CCATGCCAG CACTTATCTC GGTTACTGAG AGCAGCTCAG ATCAGAAGAC C CCT TCG               57

Pro Ser

1

GCT GTC TCG GAC TCT TGG ATG ACC TCG AAT GAG TCA GAG GAC GGG GTA               105  
Ala Val Ser Asp Ser Trp Met Thr Ser Asn Glu Ser Glu Asp Gly Val

5

10

15

TCC TCG CA   113

Ser Ser

20

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

Pro Ser Ala Val Ser Asp Ser Trp Met Thr Ser Asn Glu Ser Glu Asp

188

1

5

10

15

Gly Val Ser Ser

20

## (2) INFORMATION FOR SEQ ID NO:74:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 330 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Clone Y5-56

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..330

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

ACG TTG AAG ACA CTT TCC TCC CCT TCG GCT GTC TCG GAC TCT TGG ATG	48
Thr Leu Lys Thr Leu Ser Ser Pro Ser Ala Val Ser Asp Ser Trp Met	
1 . . . . . 5 . . . . . 10 . . . . . 15	

ACC TCG AAT GAG TCA GAG GAC GGG GTA TCC TCC TGC GAG GAG GAC ACC	96
Thr Ser Asn Glu Ser Glu Asp Gly Val Ser Ser Cys Glu Glu Asp Thr	
20 . . . . . 25 . . . . . 30	

GAC GGG GTC TTC TCA TCT GAG CTG CTC TCA GTA ACC GAG ATA AGT GCT	144
Asp Gly Val Phe Ser Ser Glu Leu Leu Ser Val Thr Glu Ile Ser Ala	
35 . . . . . 40 . . . . . 45	

GGC GAT GGA GTA CGG GGG ATG TCT TCT CCC CAT ACA GGC ATC TCT CGG	192
Gly Asp Gly Val Arg Gly Met Ser Ser Pro His Thr Gly Ile Ser Arg	

189

50	55	60	
CTA CTA CCA CAA AGA GAG GGT GTA CTG CAG TCC TCC ATG ATG ACA TCA Leu Leu Pro Gln Arg Glu Gly Val Leu Gln Ser Ser Met Met Thr Ser			240
65	70	75	80
ATG TGC GGT TCA AGA ATC CTC GCA GCA TTC TCG ATC GCT TGG AGA GCA Met Cys Gly Ser Arg Ile Leu Ala Ala Phe Ser Ile Ala Trp Arg Ala			288
85	90	95	
GCA GCC GCC GGC GGC AGA TCG GCC TCA GTC AGT TCT GAG TCT Ala Ala Ala Gly Gly Arg Ser Ala Ser Val Ser Ser Glu Ser			330
100	105	110	

## (2) INFORMATION FOR SEQ ID NO:75:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 110 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

Thr Leu Lys Thr Leu Ser Ser Pro Ser Ala Val Ser Asp Ser Trp Met			
1	5	10	15
Thr Ser Asn Glu Ser Glu Asp Gly Val Ser Ser Cys Glu Glu Asp Thr			
20	25	30	
Asp Gly Val Phe Ser Ser Glu Leu Leu Ser Val Thr Glu Ile Ser Ala			
35	40	45	
Gly Asp Gly Val Arg Gly Met Ser Ser Pro His Thr Gly Ile Ser Arg			
50	55	60	
Leu Leu Pro Gln Arg Glu Gly Val Leu Gln Ser Ser Met Met Thr Ser			
65	70	75	80
Met Cys Gly Ser Arg Ile Leu Ala Ala Phe Ser Ile Ala Trp Arg Ala			
85	90	95	

190

Ala Ala Ala Gly Gly Arg Ser Ala Ser Val Ser Ser Glu Ser  
100 105 110

**(2) INFORMATION FOR SEQ ID NO:76:**

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 195 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: both
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Clone Y5-57

(ix) FEATURE:

- (A) NAME/KEY: CDS  
(B) LOCATION: 1..195

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

```

ACG GAA AGA GCC ACG TTG AAG ACA CTT TCC TCC CCT TCG GCT GCC TCG      48
Thr Glu Arg Ala Thr Leu Lys Thr Leu Ser Ser Pro Ser Ala Ala Ser
   1           5           10          15

```

GAC TCT TGG ATG ACC TCG AAT GAG TCG GAG GAC GGG GTA TCC TCC TGC 96  
Asp Ser Trp Met Thr Ser Asn Glu Ser Glu Asp Gly Val Ser Ser Cys  
20 25 30

GAA GAG GAC ACC GAC GGG GTC TTC TCA TCT GAG CTG CTC TCA GTA ACC 144  
Glu Glu Asp Thr Asp Gly Val Phe Ser Ser Glu Leu Leu Ser Val Thr  
35 40 45

GAG ATA AGT GCT GGC GGT GGA GTA CGG GGG ATG TCT TCT CCC CAT ACG 192  
Glu Ile Ser Ala Gly Gly Gly Val Arg Gly Met Ser Ser Pro His Thr  
50 55 60

191

GGC  
Gly  
65

195

(2) INFORMATION FOR SEQ ID NO:77:

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 65 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

Thr Glu Arg Ala Thr Leu Lys Thr Leu Ser Ser Pro Ser Ala Ala Ser  
1 5 10 15

Asp Ser Trp Met Thr Ser Asn Glu Ser Glu Asp Gly Val Ser Ser Cys  
20 25 30

Glu Glu Asp Thr Asp Gly Val Phe Ser Ser Glu Leu Leu Ser Val Thr  
35 40 45

Glu Ile Ser Ala Gly Gly Gly Val Arg Gly Met Ser Ser Pro His Thr  
50 55 60

Gly  
65

**(2) INFORMATION FOR SEQ ID NO:78:**

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 115 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: both
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

192

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Clone Y5-60

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..115

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

AAG ACA CTT TCC TCC CCT TCG GCT GTC TCG GAC TCT TGG ATG ACC TCG	48
Lys Thr Leu Ser Ser Pro Ser Ala Val Ser Asp Ser Trp Met Thr Ser	
1                   5                   10                   15	

AAT GAG TCA GAG GAC GGG GTA TCC TCC TGC GAG GAG GAC ACC GAC TGG	96
Asn Glu Ser Glu Asp Gly Val Ser Ser Cys Glu Glu Asp Thr Asp Trp	
20                   25                   30	

GTC TTC TCA TCT GAG CTG C	115
Val Phe Ser Ser Glu Leu	
35	

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

Lys Thr Leu Ser Ser Pro Ser Ala Val Ser Asp Ser Trp Met Thr Ser	
1                   5                   10                   15	

Asn Glu Ser Glu Asp Gly Val Ser Ser Cys Glu Glu Asp Thr Asp Trp	
20                   25                   30	

Val Phe Ser Ser Glu Leu	
-------------------------	--

35	
----	--

## (2) INFORMATION FOR SEQ ID NO:80:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 93 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Clone Y5-63

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 19..93

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

GAGAGCAGCT CAGATGAG AAG ACA CTT TCC TCC CCT TCG GCT GTC TCG GAC	51
Lys Thr Leu Ser Ser Pro Ser Ala Val Ser Asp	
1 5 10	
TCT TGG ATG ACC TCG AAT GAG TCA GAG GAC GGG GTA TCC TCG	93
Ser Trp Met Thr Ser Asn Glu Ser Glu Asp Gly Val Ser Ser	
15 20 25	

## (2) INFORMATION FOR SEQ ID NO:81:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

194

Lys Thr Leu Ser Ser Pro Ser Ala Val Ser Asp Ser Trp Met Thr Ser  
1 5 10 15

Asn Glu Ser Glu Asp Gly Val Ser Ser  
20 25

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer Y5-10-F1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

TCAGCCATGG CTCGTGCGCC CGCGATGGTC

30

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

195

## (vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer Y5-10-R1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

CGAGGATCCA GCCGCCGGCG GCAGATC

27

## (2) INFORMATION FOR SEQ ID NO:84:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer Y5-16F1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

GATTCCATGG GTTTGGGGTT GACGGTGGCT GA

32

## (2) INFORMATION FOR SEQ ID NO:85:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

196

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer 470EP-R3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

GCGAATTCCGG ATCCCAAGGT TTCTTGCCTA GC

32

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer Y5-5-F1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

GAGGCCATGG CCTATTGTGA CAAGGTG

27

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

197

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer PGEX-R

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

GACCGTCTCC GGGAGCT

17

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 326 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Clone GE15-1

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 3..326

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

CC ATG GAG GTC TCT GAC TTC CGT GGC TCG TCT GGC TCA CCG GTC CTA  
Met Glu Val Ser Asp Phe Arg Gly Ser Ser Gly Ser Pro Val Leu  
1 5 10 15

47

TGT GAC GAA GGG CAC GCA GTA GGA ATG CTC GTG TCT GTG CTT CAC TCC  
Cys Asp Glu Gly His Ala Val Gly Met Leu Val Ser Val Leu His Ser  
20 25 30

95

198

GGT GGT AGG GTC ACC GCG GCA CGG TTC ACT AGG CCG TGG ACC CAA GTG           143  
 Gly Gly Arg Val Thr Ala Ala Arg Phe Thr Arg Pro Trp Thr Gln Val  
 35                   40                   45

CCA ACA GAT GCC AAA ACC ACC ACT GAA CCC CCT CCG GTG CCG GCC AAA           191  
 Pro Thr Asp Ala Lys Thr Thr Glu Pro Pro Pro Val Pro Ala Lys  
 50                   55                   60

GGA GTT TTC AAA GAG GCC CCG TTG TTT ATG CCT ACG GGA GCG GGA AAG           239  
 Gly Val Phe Lys Glu Ala Pro Leu Phe Met Pro Thr Gly Ala Gly Lys  
 65                   70                   75

AGC ACT CGC GTC CCG TTG GAG TAC GCC AAC ATG GGG CAC AAG GTC TTA           287  
 Ser Thr Arg Val Pro Leu Glu Tyr Gly Asn Met Gly His Lys Val Leu  
 80                   85                   90                   95

ATC TTG AAC CCC TCA GTG GCC ACT GTG CCG GCG ATG GGC                   326  
 Ile Leu Asn Pro Ser Val Ala Thr Val Arg Ala Met Gly  
 100                   105

## (2) INFORMATION FOR SEQ ID NO:89:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 108 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

Met Glu Val Ser Asp Phe Arg Gly Ser Ser Gly Ser Pro Val Leu Cys  
 1                   5                   10                   15

Asp Glu Gly His Ala Val Gly Met Leu Val Ser Val Leu His Ser Gly  
 20                   25                   30

Gly Arg Val Thr Ala Ala Arg Phe Thr Arg Pro Trp Thr Gln Val Pro  
 35                   40                   45

Thr Asp Ala Lys Thr Thr Glu Pro Pro Pro Val Pro Ala Lys Gly  
 50                   55                   60

199

Val Phe Lys Glu Ala Pro Leu Phe Met Pro Thr Gly Ala Gly Lys Ser			
65	70	75	80
Thr Arg Val Pro Leu Glu Tyr Gly Asn Met Gly His Lys Val Leu Ile			
85	90	95	
Leu Asn Pro Ser Val Ala Thr Val Arg Ala Met Gly			
100	105		

## (2) INFORMATION FOR SEQ ID NO:90:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 138 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Clone GE17-2

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..138

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

GGT GAT GAG GTT CTC ATC GGC GTC TTC CAG GAT GTG AAT CAT TTG CCT	48
Gly Asp Glu Val Leu Ile Gly Val Phe Gln Asp Val Asn His Leu Pro	
1 5 10 15	
CCC GGG TTT GTT CCG ACC GCG CCT GTT GTC ATC CGA CGG TGC GGA AAG	96
Pro Gly Phe Val Pro Thr Ala Pro Val Val Ile Arg Arg Cys Gly Lys	
20 25 30	
GCC TTC TTG GGG GTC ACA AAG GCT GCC TTG ACA GGT CGG GAT	138
Gly Phe Leu Gly Val Thr Lys Ala Ala Leu Thr Gly Arg Asp	

200

35

40

45

## (2) INFORMATION FOR SEQ ID NO:91:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

Gly Asp Glu Val Leu Ile Gly Val Phe Gln Asp Val Asn His Leu Pro  
1 5 10 15

Pro Gly Phe Val Pro Thr Ala Pro Val Val Ile Arg Arg Cys Gly Lys  
20 25 30

Gly Phe Leu Gly Val Thr Lys Ala Ala Leu Thr Gly Arg Asp  
35 40 45

## (2) INFORMATION FOR SEQ ID NO:92:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA to mRNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Primer GE15F

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

201

GCCGCCATGG AGGTCTCTGA CTTCCGTG

28

## (2) INFORMATION FOR SEQ ID NO:93:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer GE15R

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

CGCGGGATCC GCCCATCGCC CGCACAGTGG C

31

## (2) INFORMATION FOR SEQ ID NO:94:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer GE17F

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

202

CGCTCCATGG GTGATGAGGT TCTCATCGGC G

31

## (2) INFORMATION FOR SEQ ID NO:95:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Primer GE17R

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

GTAAGTCAGG ATCCCGACCT GTCAAGGC

28

## (2) INFORMATION FOR SEQ ID NO:96:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 452 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: NcoI/EcoRI-containing fragment of pGEX-HISb-GE3-s HGV plasmid

203

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

CRAAATCGGA TCTGGTTCCG CGTGGTTCCA TGGTCTCATG GGACGGGAC GCTCGTGCGC	60
CCCGCGATGGT CTATGGCCCT GGGCAAAGTG TTACCATTGA CGGGGAGCGC TACACCTTGC	120
CTCATCAACT GAGGCTCAGG AATGTGGCAC CCTCTGAGGT TTCATCCGAG GTGTCCATTG	180
ACATTGGGAC GGAGACTGAA GACTCAGAAC TGACTGAGGC CGATCTGCCG CCGGCGGCTG	240
CTGCTCTCCA AGCGATCGAG AATGCTGCCA GGATTCTTGA ACCGCACATT GATGTCATCA	300
TGGAGGACTG CAGTACACCC TCTCTTGTG CTAGTAGCCG AGAGATGCCT GTATGGGAG	360
AAGACATCCC CCGTACTCCA TCGCCAGCAC TTATCGGATC CCACCATCAC CATCACCATT	420
AGAATTTCATC GTGACTGACT GACGATCTAC CT	452

## (2) INFORMATION FOR SEQ ID NO:97:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Primer 470EP-F8

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

GCTGAATTG CCATGGCGAC GTGCGCATTC AGGGGTGGA	39
---	----

## (2) INFORMATION FOR SEQ ID NO:98:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid

204

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer 470EP-F9

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

GCTAGATCTG GCAACATGGG GCACAAGGTC

30

(2) INFORMATION FOR SEQ ID NO:98:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer 470EP-R9

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

CACAGATCTC GCGTAGTAGT AGCGTCCAGA

30

(2) INFORMATION FOR SEQ ID NO:100:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid

205

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer 9E3-REV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

GCTGGCTGAG GCACGGTTGG TC

22

(2) INFORMATION FOR SEQ ID NO:101:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer E39-94PR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

CACCATCATC ACAGCATCTG GC

22

(2) INFORMATION FOR SEQ ID NO:102:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 32 base pairs
  - (B) TYPE: nucleic acid

206

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer GEP-F12

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

GCAACCATGG AACCTGCCAA ACCCCTGACC TT

32

(2) INFORMATION FOR SEQ ID NO:103:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer GEP-F14

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

TTGGGATCCC TCGTGTTCGG CCATTCTAAG

30

(2) INFORMATION FOR SEQ ID NO:104:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid

207

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer GEP-F15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

CGGGCCATGG TGCCCTTCGT CAATAGGACA

30

(2) INFORMATION FOR SEQ ID NO:105:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer GEP-R16

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

TGCGAATCCT CGGCCCTGGT TGCCCAG

27

(2) INFORMATION FOR SEQ ID NO:106:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid

208

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer GEP-R12

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

AGCCCCATGG AAGGTCGTGA A

21

(2) INFORMATION FOR SEQ ID NO:107:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer GEP-R13

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

TATGGATCCT GGTAAATCAT TGCCCCACCT

30

(2) INFORMATION FOR SEQ ID NO:108:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 27 base pairs
  - (B) TYPE: nucleic acid

209

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer GEP-R14

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

GGAGGATCCG CGACCCGCCA CGGAAGT

27

(2) INFORMATION FOR SEQ ID NO:109:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 29 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer GEP-R15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

CTTGCCATGG CCAGCTGGTT CACCCACCA

29

(2) INFORMATION FOR SEQ ID NO:110:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid

210

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Primer GEP-F17

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

GCAGGGATCCC CTCTGGAAGG TCCCATTGAA

30

(2) INFORMATION FOR SEQ ID NO:111:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 138 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: both
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Reverse-Frame Antigen K1-2-3A

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..138

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

AGC CTT AGA ATG GCA GAA CAC GAG GTG CCT TCC GGT TCG CAT CCG CTC  
Ser Leu Arg Met Ala Glu His Glu Val Pro Ser Gly Ser His Pro Leu

1

5

10

15

48

211

GAG GGG TAT TCC ATG TCC ATA AAA GGG AAT CTC GCC CAC GTC CAA TTT 96  
 Glu Gly Tyr Ser Met Ser Ile Lys Gly Asn Leu Ala His Val Gln Phe  
           20                 25                 30

TGT CTC AAT TAT GGA AGG GTG CTG CGT CAT AGG GGA GAA TTC      138  
 Cys Leu Asn Tyr Gly Arg Val Leu Arg His Arg Gly Glu Phe  
               35                  40                  45

**(2) INFORMATION FOR SEQ ID NO:112:**

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 46 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

Ser Leu Arg Met Ala Glu His Glu Val Pro Ser Gly Ser His Pro Leu  
1 5 10 15

Glu Gly Tyr Ser Met Ser Ile Lys Gly Asn Leu Ala His Val Gln Phe  
20 25 30

Cys Leu Asn Tyr Gly Arg Val Leu Arg His Arg Gly Glu Phe  
35 40 45

(2) INFORMATION FOR SEQ ID NO:113:

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 240 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: both
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

212

## (C) INDIVIDUAL ISOLATE: Reverse-Frame Antigen K3-10-1D

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..240

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

CAC CAT CAT CAC AGC ATC TGG CCA GAT GTA CGA GGC CAA GCA CCA GGC	48
His His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Gly	
1 5 10 15	

AAA GGT CAG GGG TTT GGC AGG CCG CCC CTC CCC GAG GGG GCT CCT CAC	96
Lys Gly Gln Gly Phe Gly Arg Pro Pro Leu Pro Glu Gly Ala Pro His	
20 25 30	

CAC CCT TTG ACG GAT CGC CTG ATA CCC CTT ACA CCA CGT CCT ATA GAT	144
His Pro Leu Thr Asp Arg Leu Ile Pro Leu Thr Pro Arg Pro Ile Asp	
35 40 45	

CAC GGC TTT GTG CCT CCA CCC CCC GCG CTC ATC GAG CTC AGG AGC GCA	192
His Gly Phe Val Pro Pro Pro Ala Leu Ile Glu Leu Arg Ser Ala	
50 55 60	

ATG GCC CAA GCC ATC ACA CAG CAC TCG ACC AAC CGT GCC TCA GCC AGC	240
Met Ala Gln Ala Ile Thr Gln His Ser Thr Asn Arg Ala Ser Ala Ser	
65 70 75 80	

## (2) INFORMATION FOR SEQ ID NO:114:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 80 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

His His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Gly	
1 5 10 15	

213

Lys Gly Gln Gly Phe Gly Arg Pro Pro Leu Pro Glu Gly Ala Pro His  
 20 25 30

His Pro Leu Thr Asp Arg Leu Ile Pro Leu Thr Pro Arg Pro Ile Asp  
 35 40 45

His Gly Phe Val Pro Pro Pro Ala Leu Ile Glu Leu Arg Ser Ala  
 50 55 60

Met Ala Gln Ala Ile Thr Gln His Ser Thr Asn Arg Ala Ser Ala Ser  
 65 70 75 80

**(2) INFORMATION FOR SEQ ID NO:115:**

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 318 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

**(ii) MOLECULE TYPE:** cDNA to mRNA

**(iii) HYPOTHETICAL:** NO

**(iv) ANTI-SENSE:** NO

**(vi) ORIGINAL SOURCE:**

**(C) INDIVIDUAL ISOLATE:** Reverse-Frame Antigen K3-11-1A

**(ix) FEATURE:**

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..318

**(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:**

CAC CAT CAT CAC AGC ATC TGG CCA GAT GTA CGA GGC CAA GCA CCA GGC 48  
 His His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Gly  
 1 5 10 15

AAA GGT CAG GGG TTT GGC AGG CCG CCC CTC CCC GAG GGG GCT CCT CAC 96  
 Lys Gly Gln Gly Phe Gly Arg Pro Pro Leu Pro Glu Gly Ala Pro His  
 20 25 30

214

CAC CCT TTG ACC GAT TGC CTG GTA CCC CTT ACA CCA CGT CCT ATA GAT	144		
His Pro L u Thr Asp Cys Leu Val Pro Leu Thr Pro Arg Pro Ile Asp			
35	40	45	
CAC GGC TTT GTG CCT CCA CCC CCT GCC CTC ATC GAG CTC AGG AGC GCA	192		
His Gly Phe Val Pro Pro Pro Ala Leu Ile Glu Leu Arg Ser Ala			
50	55	60	
ATG GCC CAA GCT ACC ACA CTG GCC ACC ACC CAG CCC AAC ACC GAA GTG	240		
Met Ala Gln Ala Thr Thr Leu Ala Thr Thr Gln Pro Asn Thr Glu Val			
65	70	75	80
TCC ACC TCG AAT GTA GCA TCG AAG CAG AAC TCG GCC CCG AGC ACT GAG	288		
Ser Thr Ser Asn Val Ala Ser Lys Gln Asn Ser Ala Pro Ser Thr Glu			
85	90	95	
GTG CGC CCG CGG GTC GCC GAA ATC CCC ATC	318		
Val Arg Pro Arg Val Ala Glu Ile Pro Ile			
100	105		

(2) INFORMATION FOR SEQ ID NO:116:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 106 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

His His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Gly  
1 5 10 15

Lys Gly Gln Gly Phe Gly Arg Pro Pro Leu Pro Glu Gly Ala Pro His  
                   20                  25                  30

His Pro Leu Thr Asp Cys Leu Val Pro Leu Thr Pro Arg Pro Ile Asp  
35 40 . 45

His Gly Phe Val Pro Pro Pro Pro Ala Leu Ile Glu Leu Arg Ser Ala  
50 55 60

215

Met Ala Gln Ala Thr Thr Leu Ala Thr Thr Gln Pro Asn Thr Glu Val  
 65                    70                    75                    80

Ser Thr Ser Asn Val Ala Ser Lys Gln Asn Ser Ala Pro Ser Thr Glu  
 85                    90                    95

Val Arg Pro Arg Val Ala Glu Ile Pro Ile  
 100                  105

## (2) INFORMATION FOR SEQ ID NO:117:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 240 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Reverse-Frame Antigen K3-14-2A

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..240

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

CAC CAT CAT CAC AGC ATC TGG CCA GAC GTA CGA GGC CAA GCA CCA GGC        48  
 His His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Gly  
 1                    5                    10                    15

AAA GGT CAG GGG TTT GGC AGG CCG CCC CTC CCC GAG GGG GCT CCT CAC        96  
 Lys Gly Gln Gly Phe Gly Arg Pro Pro Leu Pro Glu Gly Ala Pro His  
 20                  25                  30

CAC CCT TTG ACG GAT TGC CTG GTA CCC CTT ACA CCA CGT CCT ATA GAT        144  
 His Pro Leu Thr Asp Cys Leu Val Pro Leu Thr Pro Arg Pro Ile Asp  
 35                  40                  45

216

CAC	GGC	TTT	GTG	CCT	CCA	CCC	CCT	GCG	CTC	ATC	GAG	CTC	AGG	AGC	GCA	192
His	Gly	Phe	Val	Pro	Pro	Pro	Pro	Ala	Leu	Ile	Glu	Leu	Arg	Ser	Ala	
50				55					60							
ATG	GCC	CAA	GCT	ACC	ACA	CAG	CAC	TCG	ACC	AAC	CGT	GCC	TCA	GCC	AGC	240
Met	Ala	Gln	Ala	Thr	Thr	Gln	His	Ser	Thr	Asn	Arg	Ala	Ser	Ala	Ser	
65				70					75			80				

## (2) INFORMATION FOR SEQ ID NO:118:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 80 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

His	His	His	His	Ser	Ile	Trp	Pro	Asp	Val	Arg	Gly	Gln	Ala	Pro	Gly	
1					5				10				15			

Lys	Gly	Gln	Gly	Phe	Gly	Arg	Pro	Pro	Leu	Pro	Glu	Gly	Ala	Pro	His	
					20			25			30					

His	Pro	Leu	Thr	Asp	Cys	Leu	Val	Pro	Leu	Thr	Pro	Arg	Pro	Ile	Asp	
					35			40			45					

His	Gly	Phe	Val	Pro	Pro	Pro	Ala	Leu	Ile	Glu	Leu	Arg	Ser	Ala	
					50			55			60				

Met	Ala	Gln	Ala	Thr	Thr	Gln	His	Ser	Thr	Asn	Arg	Ala	Ser	Ala	Ser	
					65			70			75		80			

## (2) INFORMATION FOR SEQ ID NO:119:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 240 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

217

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Reverse-Frame Antigen K3-14-3A

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..240

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

CAC CAT CAT CAC AGC ATC TGG CCA GAT GTA CGA GGC CAA GCA CCA GGC	48	
His His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Gly		
1                   5                   10                   15		
AAA GGT CAG GGG TTT GGC AGG CCG CCC CTC CCC GAG GGG GCT CCT CAC	96	
Lys Gly Gln Gly Phe Gly Arg Pro Pro Leu Pro Glu Gly Ala Pro His		
20                   25                   30		
CAC CCT TTG ACG GAT TGC CTG GTA CCC CTT ACA CCA CGT CCT ATA GAT	144	
His Pro Leu Thr Asp Cys Leu Val Pro Leu Thr Pro Arg Pro Ile Asp		
35                   40                   45		
CAC GGC TTT GTG CCT CCA CCC CCT GCG CTC ATC GAG CTC AGG AGC GCA	192	
His Gly Phe Val Pro Pro Pro Ala Leu Ile Glu Leu Arg Ser Ala		
50                   55                   60		
ATG GCC CAA GCT ACC ACA CAG CAC TCG ACC AAC CGT GCC TCA GCC AGC	240	
Met Ala Gln Ala Thr Thr Gln His Ser Thr Asn Arg Ala Ser Ala Ser		
65                   70                   75                   80		

(2) INFORMATION FOR SEQ ID NO:120:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 80 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

218

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Gly  
1 5 10 15

Lys Gly Gln Gly Phe Gly Arg Pro Pro Leu Pro Glu Gly Ala Pro His  
20 25 30

His Pro Leu Thr Asp Cys Leu Val Pro Leu Thr Pro Arg Pro Ile Asp  
35 40 45

His Gly Phe Val Pro Pro Pro Ala Leu Ile Glu Leu Arg Ser Ala  
50 55 60

Met Ala Gln Ala Thr Thr Gln His Ser Thr Asn Arg Ala Ser Ala Ser  
65 70 75 80

(2) INFORMATION FOR SEQ ID NO:121:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 240 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Reverse-Frame Antigen K3-14-5A

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..240

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

219

CAC CAT CAT CAC ACC ATC TGG CCA GAT GTA CGA GGC CAA GCA CCG AGC	48
His His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Ser	
1 5 10 15	
AAA GGT CAG GGG TTT GGC AGG CCG CCC CTC CCC GAG GGG GCT CCT CAC	96
Lys Gly Gln Gly Phe Gly Arg Pro Pro Leu Pro Glu Gly Ala Pro His	
20 25 30	
CAC CCT TTG ACG GAT TGC CTG GTA CCC CTT ACA CCA CGT CCT ATA GAT	144
His Pro Leu Thr Asp Cys Leu Val Pro Leu Thr Pro Arg Pro Ile Asp	
35 40 45	
CAC GGC TTT GTG CCT CAC CCC CCT GCG CTC ATC GAG CTC AGG AGC GCA	192
His Gly Phe Val Pro His Pro Pro Ala Leu Ile Glu Leu Arg Ser Ala	
50 55 60	
ATG GCC CAA GCC ATC ACA CAG CAC TCG ACC AAC CGT GCC TCA GCC AGC	240
Met Ala Gln Ala Ile Thr Gln His Ser Thr Asn Arg Ala Ser Ala Ser	
65 70 75 80	

## (2) INFORMATION FOR SEQ ID NO:122:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 80 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

His His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Ser	
1 5 10 15	
Lys Gly Gln Gly Phe Gly Arg Pro Pro Leu Pro Glu Gly Ala Pro His	
20 25 30	
His Pro Leu Thr Asp Cys Leu Val Pro Leu Thr Pro Arg Pro Ile Asp	
35 40 45	
His Gly Phe Val Pro His Pro Pro Ala Leu Ile Glu Leu Arg Ser Ala	
50 55 60	

220

Met Ala Gln Ala Ile Thr Gln His Ser Thr Asn Arg Ala Ser  
 65                   70                   75                   80

## (2) INFORMATION FOR SEQ ID NO:123:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 240 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Reverse-Frame Antigen K3-14-6A

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..240

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

CAC CAT CAT CAC AGC ATC TGG CCA GAT GTA CGA GGC CAA GCA CCA GGC	48
His His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Gly	
1               5                   10                   15	

AAA GGT CAG GGG TTT GGC GGG CCG CCC CTC CCC GAG GGG GCT CCT CAC	96
Lys Gly Gln Gly Phe Gly Pro Pro Leu Pro Glu Gly Ala Pro His	
20              25                   30	

CAC CCT TTG ACG GAT TGC CTG GTA CCC CTT ACA CCA CGT CCT ATA GAT	144
His Pro Leu Thr Asp Cys Leu Val Pro Leu Thr Pro Arg Pro Ile Asp	
35              40                   45	

CAC GGC TTT GTG CCT CCA CCC CCC GCG CTC ATC GAG CTC AGG AGC GCA	192
His Gly Phe Val Pro Pro Pro Ala Leu Ile Glu Leu Arg Ser Ala	
50              55                   60	

221

ATG GCC CAA GCT ACC ACA CAG CAC TCG ACC AAC CGT GCC TCA GCC AGC           240  
Met Ala Gln Ala Thr Thr Gln His Ser Thr Asn Arg Ala Ser Ala Ser  
65                   70                   75                   80

## (2) INFORMATION FOR SEQ ID NO:124:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 80 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

His His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Gly  
1               5                   10                   15

Lys Gly Gln Gly Phe Gly Gly Pro Pro Leu Pro Glu Gly Ala Pro His  
20               25                   30

His Pro Leu Thr Asp Cys Leu Val Pro Leu Thr Pro Arg Pro Ile Asp  
35               40                   45

His Gly Phe Val Pro Pro Pro Ala Leu Ile Glu Leu Arg Ser Ala  
50               55                   60

Met Ala Gln Ala Thr Thr Gln His Ser Thr Asn Arg Ala Ser Ala Ser  
65               70                   75                   80

## (2) INFORMATION FOR SEQ ID NO:125:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 243 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA to mRNA

## (iii) HYPOTHETICAL: NO

222

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Reverse-Frame Antigen K3-17-1A

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..243

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

CAC CAT CAT CAC AGC ATC TGG CCA GAT GTA CGA GGC CAA GCA CCA GGC	48	
His His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Gly		
1                   5                   10                   15		
AAA GGT CAG GGG TTT GGC AGG CCG CCC CTC CCC GAG GGG GCT CCT CAC	96	
Lys Gly Gln Gly Phe Gly Arg Pro Pro Leu Pro Glu Gly Ala Pro His		
20                   25                   30		
CAC CCT TTG ACG GAT TGC CTG GTA CCC CTT ACA CCA CGT CCT ATA GAT	144	
His Pro Leu Thr Asp Cys Leu Val Pro Leu Thr Pro Arg Pro Ile Asp		
35                   40                   45		
CAC GGC TTT GTG CCT CCA CCC CCC CCT GCG CTC ATC GAG CTC AGG AGC	192	
His Gly Phe Val Pro Pro Pro Pro Ala Leu Ile Glu Leu Arg Ser		
50                   55                   60		
GCA ATG GCC CAA GCT ACC ACA CAG CAC TCG ACC AAC CGT GCC TCA GCC	240	
Ala Met Ala Gln Ala Thr Thr Gln His Ser Thr Asn Arg Ala Ser Ala		
65                   70                   75                   80		
AGC	243	
Ser		

(2) INFORMATION FOR SEQ ID NO:126:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 81 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

223

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Gly  
1               5                           10                           15

Lys Gly Gln Gly Phe Gly Arg Pro Pro Leu Pro Glu Gly Ala Pro His  
20   25                           30

His Pro Leu Thr Asp Cys Leu Val Pro Leu Thr Pro Arg Pro Ile Asp  
35    40                           45

His Gly Phe Val Pro Pro Pro Pro Ala Leu Ile Glu Leu Arg Ser  
50    55                           60

Ala Met Ala Gln Ala Thr Thr Gln His Ser Thr Asn Arg Ala Ser Ala  
65    70                           75                           80

Ser

(2) INFORMATION FOR SEQ ID NO:127:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 156 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Reverse-Frame Antigen K3-8-3A

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..156

224

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

CAC CAT CAT CAC AGC ATC TGG CCA GAT GTA CGA GGC CAA GCA CCA GGC	48
His His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Gly	
1                   5                   10                   15	
AAA GGT CAG GGG TTT GCC AGG CCG CCC CTC CCC CCT GCG CTC ATC GAG	96
Lys Gly Gln Gly Phe Gly Arg Pro Pro Leu Pro Pro Ala Leu Ile Glu	
20                   25                   30	
CTC AGG AGC GCA ATG GCC CAA GCC ATC ACA CAG CAC TCG ACC AAC CGT	144
Leu Arg Ser Ala Met Ala Gln Ala Ile Thr Gln His Ser Thr Asn Arg	
35                   40                   45	
GCC TCA GCC AGC	156
Ala Ser Ala Ser	
50	

## (2) INFORMATION FOR SEQ ID NO:128:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 52 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

His His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Gly	
1                   5                   10                   15	
Lys Gly Gln Gly Phe Gly Arg Pro Pro Leu Pro Pro Ala Leu Ile Glu	
20                   25                   30	
Leu Arg Ser Ala Met Ala Gln Ala Ile Thr Gln His Ser Thr Asn Arg	
35                   40                   45	
Ala Ser Ala Ser	
50	

## (2) INFORMATION FOR SEQ ID NO:129:

225

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 240 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Reverse-Frame Antigen K3-8-4C

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..240

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

CAC CAT CAT CAC AGC ATC TGG CCA GAT GTA CGA GGC CAA GCA CCA GGC	48
His His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Gly	
1                   5                   10                   15	
AAA GGT CAG GGG TTT GGC AGG CCG CCC CTC CCC GAG GGG GCT CCT CAC	96
Lys Gly Gln Gly Phe Gly Arg Pro Pro Leu Pro Glu Gly Ala Pro His	
20                   25                   30	
CAC CCT TTG ACG GAT TGC CTG GTA CCC CTT ACA TCA CGT CCT ATA GAT	144
His Pro Leu Thr Asp Cys Leu Val Pro Leu Thr Ser Arg Pro Ile Asp	
35                   40                   45	
CAC GGC TTT GTG CCT CCA CCC CCT GCG CTC ATC GAG CTC AGG AGC GCA	192
His Gly Phe Val Pro Pro Pro Ala Leu Ile Glu Leu Arg Ser Ala	
50                   55                   60	
ATG GCC CAA GCC ATC ACA CAG CAC TCG ACC AAC CGT GCC TCA GCC AGC	240
Met Ala Gln Ala Ile Thr Gln His Ser Thr Asn Arg Ala Ser Ala Ser	
65                   70                   75                   80	

(2) INFORMATION FOR SEQ ID NO:130:

226

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 80 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

His	His	His	His	Ser	Ile	Trp	Pro	Asp	Val	Arg	Gly	Gln	Ala	Pro	Gly
1					5				10					15	

Lys	Gly	Gln	Gly	Phe	Gly	Arg	Pro	Pro	Leu	Pro	Glu	Gly	Ala	Pro	His
									25					30	

His	Pro	Leu	Thr	Asp	Cys	Leu	Val	Pro	Leu	Thr	Ser	Arg	Pro	Ile	Asp
					35				40					45	

His	Gly	Phe	Val	Pro	Pro	Pro	Pro	Ala	Leu	Ile	Glu	Leu	Arg	Ser	Ala
					50				55					60	

Met	Ala	Gln	Ala	Ile	Thr	Gln	His	Ser	Thr	Asn	Arg	Ala	Ser	Ala	Ser
					65				70					75	80

## (2) INFORMATION FOR SEQ ID NO:131:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 239 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA to mRNA

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Reverse-Frame Antigen K3-8-5A

## (ix) FEATURE:

- (A) NAME/KEY: CDS

227

(B) LOCATION: 1..239

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

CAC CAT CAT CAC AGC ATC TGG CCA GAT GTA CGA GGC CAA GCA CCA GGC	48
His His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Gly	
1 5 10 15	
AAA GGT CAG GGG TTT GGC AGG CCG CCC CTC CCC GAG GGG GCT CCT CAC	96
Lys Gly Gln Gly Phe Gly Arg Pro Pro Leu Pro Glu Gly Ala Pro His	
20 25 30	
CAC CCT TTG ACG GAT TGC CTG GTA CCC CTT ACA CCA CGT CCT ATA GAT	144
His Pro Leu Thr Asp Cys Leu Val Pro Leu Thr Pro Arg Pro Ile Asp	
35 40 45	
CAC GGC TTA GTG CCT CCA CCC CCT GCG CTC ATC GAG CTC AGG AGC GCA	192
His Gly Leu Val Pro Pro Pro Ala Leu Ile Glu Leu Arg Ser Ala	
50 55 60	
ATG GCC CAA GCT ACC ACA CAG CAC TCG ACC AAC CGT GCC TCA GCC AG	239
Met Ala Gln Ala Thr Thr Gln His Ser Thr Asn Arg Ala Ser Ala	
65 70 75	

(2) INFORMATION FOR SEQ ID NO:132:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 79 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

His His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Gly	
1 5 10 15	
Lys Gly Gln Gly Phe Gly Arg Pro Pro Leu Pro Glu Gly Ala Pro His	
20 25 30	
His Pro Leu Thr Asp Cys Leu Val Pro Leu Thr Pro Arg Pro Ile Asp	

228

35

40

45

His Gly Leu Val Pro Pro Pro Ala Leu Ile lu Leu Arg Ser Ala  
 50 55 60

Met Ala Gln Ala Thr Thr Gln His Ser Thr Asn Arg Ala Ser Ala  
 65 70 75

## (2) INFORMATION FOR SEQ ID NO:133:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 427 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Reverse-Frame Antigen K3-8-6A

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..427

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

CAC CAC CCT TTG ACG GAT TGC CTG GTA CCC CTT ACA CCA CGT CCT ATA 48  
 His His Pro Leu Thr Asp Cys Leu Val Pro Leu Thr Pro Arg Pro Ile  
 1 5 10 15

GAT CAC GGC TTT GTG CCT CCA CCC CCT GCG CTC ATC GAG CTC AGG AGC 96  
 Asp His Gly Phe Val Pro Pro Pro Ala Leu Ile Glu Leu Arg Ser  
 20 25 30

GCA ATG GCC CAA GCT ACC ACA CTG GCC ACC ACC CAG CCC AAC ACC GAA 144  
 Ala Met Ala Gln Ala Thr Thr Leu Ala Thr Thr Gln Pro Asn Thr Glu  
 35 40 45

229

GTG TCC ACC TCG AAT GTA GCA TCG AAG CAG AAC TCG ACC CCG AGC ACT		192
Val Ser Thr Ser Asn Val Ala Ser Lys Ile Asn Ser Thr Pro Ser Thr		
50	55	60
GAG GTG CGC CCG CGG GTC GCC GAA ATC CCC ATC AAG AGG GCC AGC GGG		240
Glu Val Arg Pro Arg Val Ala Glu Ile Pro Ile Lys Arg Ala Ser Gly		
65	70	75
AAA GCT CCC CGA GCA AGC TTC CAC AAC ACG AGG AAC ATC AGG CGT TGG		288
Lys Ala Pro Arg Ala Ser Phe His Asn Thr Arg Asn Ile Arg Arg Trp		
85	90	95
GGT CCC AAC CAT CTA AAG TAC AGC ACT AGG TTT GCC AAA CCC AAT ATC		336
Gly Pro Asn His Leu Lys Tyr Ser Thr Arg Phe Ala Lys Pro Asn Ile		
100	105	110
ATA CTG ACG ACC GGG AGT CCC AGA CAC CAG GAC AGG GCA GGG CCC GCG		384
Ile Leu Thr Thr Gly Ser Pro Arg His Gln Asp Arg Ala Gly Pro Ala		
115	120	125
GAG ACC TCA CCT GCC GCG GCT TCC ACA GCC GGC AGC CCT A		427
Glu Thr Ser Pro Ala Ala Ala Ser Thr Ala Gly Ser Pro		
130	135	140

## (2) INFORMATION FOR SEQ ID NO:134:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 142 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

His His Pro Leu Thr Asp Cys Leu Val Pro Leu Thr Pro Arg Pro Ile		
1	5	10
Asp His Gly Phe Val Pro Pro Pro Ala Leu Ile Glu Leu Arg Ser		
20	25	30
Ala Met Ala Gln Ala Thr Thr Leu Ala Thr Thr Gln Pro Asn Thr Glu		
35	40	45

230

Val Ser Thr Ser Asn Val Ala Ser Lys Gln Asn Ser Thr Pro Ser Thr  
 50                    55                    60

Glu Val Arg Pro Arg Val Ala Glu Ile Pro Ile Lys Arg Ala Ser Gly  
 65                    70                    75                    80

Lys Ala Pro Arg Ala Ser Phe His Asn Thr Arg Asn Ile Arg Arg Trp  
 85                    90                    95

Gly Pro Asn His Leu Lys Tyr Ser Thr Arg Phe Ala Lys Pro Asn Ile  
 100                  105                  110

Ile Leu Thr Thr Gly Ser Pro Arg His Gln Asp Arg Ala Gly Pro Ala  
 115                  120                  125

Glu Thr Ser Pro Ala Ala Ala Ser Thr Ala Gly Ser Pro  
 130                  135                  140

**(2) INFORMATION FOR SEQ ID NO:135:**

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 240 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

**(ii) MOLECULE TYPE:** cDNA to mRNA

**(iii) HYPOTHETICAL:** NO

**(iv) ANTI-SENSE:** NO

**(vi) ORIGINAL SOURCE:**

(C) INDIVIDUAL ISOLATE: Reverse-Frame Antigen K3-8-7C

**(ix) FEATURE:**

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..240

**(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:**

CAC CAT CAT CAC AGC ATC TGG CCA GAT GTA CGA GGC CAA GCA CCA GGC  
 His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Gly

231

1

5

10

15

AAA GGT CAG GGG TTT GGC AGG CCG CCC CTC CCC GAG GGG GCT CCT CAC  
 Lys Gly Gln Gly Phe Gly Arg Pro Pro Leu Pro Glu Gly Ala Pro His

20

25

30

96

CAC CCT CTG ACG GAT TGC CTG GTA CCC CTT ACA CCA CGT CCT ATA GAT  
 His Pro Leu Thr Asp Cys Leu Val Pro Leu Thr Pro Arg Pro Ile Asp

35

40

45

144

CAC GCC TTT GTG CCT CCA CCC CCT GCG CTC ATC GAG CTC AGG AGC GCA  
 His Gly Phe Val Pro Pro Pro Ala Leu Ile Glu Leu Arg Ser Ala

50

55

60

192

ACG GCC CAA GCC ATC ACA CAG CAC TCG ACC AAC CGT GCC TCA GCC AGC  
 Thr Ala Gln Ala Ile Thr Gln His Ser Thr Asn Arg Ala Ser Ala Ser

65

70

75

80

240

## (2) INFORMATION FOR SEQ ID NO:136:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 80 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

His His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Gly  
 1 5 10 15

Lys Gly Gln Gly Phe Gly Arg Pro Pro Leu Pro Glu Gly Ala Pro His  
 20 25 30

His Pro Leu Thr Asp Cys Leu Val Pro Leu Thr Pro Arg Pro Ile Asp  
 35 40 45

His Gly Phe Val Pro Pro Pro Ala Leu Ile Glu Leu Arg Ser Ala  
 50 55 60

Thr Ala Gln Ala Ile Thr Gln His Ser Thr Asn Arg Ala Ser Ala Ser  
 65 70 75 80

## (2) INFORMATION FOR SEQ ID NO:137:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 235 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Reverse-Frame Antigen Y10-13-1

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..235

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

GTC AGC CGA GGC CCC ACG CCG CAC CGA TGG AAT GGG AAC CTA ACC GAC	48
Val Ser Arg Gly Pro Thr Pro His Arg Trp Asn Gly Asn Leu Thr Asp	
1 5 10 15	
CCG GTC TCG GGT CAG CAG TCC CTC ACA CAG GTG CCG CGG GAG GCA GGC	96
Pro Val Ser Gly Gln Gln Ser Leu Thr Gln Val Pro Arg Glu Ala Gly	
20 25 30	
CGA CGG TCC AGA ACA CAC GTC ACG CAC GGG ATT CTC CAC TCG GAG AGC	144
Arg Arg Ser Arg Thr His Val Thr His Gly Ile Leu His Ser Glu Ser	
35 40 45	
CCA GGC ACC GTG TCG CGA TCC GAT GAT CCA AGT GCG GCT ATG GTG CAG	192
Pro Gly Thr Val Ser Arg Ser Asp Asp Pro Ser Ala Ala Met Val Gln	
50 55 60	
GTG GCA GAG CCA ACC GGC ACT AAA CTC CAC ACA TCT ATC TTC G	235
Val Ala Glu Pro Thr Gly Thr Lys Leu His Thr Ser Ile Phe	
65 70 75	

## (2) INFORMATION FOR SEQ ID NO:138:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 78 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

Val Ser Arg Gly Pro Thr Pro His Arg Trp Asn Gly Asn Leu Thr Asp  
1 5 10 15

Pro Val Ser Gly Gln Gln Ser Leu Thr Gln Val Pro Arg Glu Ala Gly  
20 25 30

Arg Arg Ser Arg Thr His Val Thr His Gly Ile Leu His Ser Glu Ser  
35 40 45

Pro Gly Thr Val Ser Arg Ser Asp Asp Pro Ser Ala Ala Met Val Gln  
50 55 60

Val Ala Glu Pro Thr Gly Thr Lys Leu His Thr Ser Ile Phe  
65 70 75

## (2) INFORMATION FOR SEQ ID NO:139:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 181 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL IS LATE: Reverse-Frame Antigen Y10-13-2

234

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..181

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

TCG GGC CAG CAG TCC CTC ACA CAG GTG CCG CAG GAG GCA GGC CGA CGG	48	
Ser Gly Gln Gln Ser Leu Thr Gln Val Pro Gln Glu Ala Gly Arg Arg		
1                   5                   10                   15		
TCC AGA ACA CAC GTC ACG CAC GGG ATT CCC CAC TCG GAG AGC TCA GGC	96	
Ser Arg Thr His Val Thr His Gly Ile Pro His Ser Glu Ser Ser Gly		
20                  25                  30		
ACC GTG TCG CGA TCC GAT GAT CCA AGT GCG GTT ATG GTG CAG GTG GCA	144	
Thr Val Ser Arg Ser Asp Asp Pro Ser Ala Val Met Val Gln Val Ala		
35                  40                  45		
GAG CCA ACT GGC ACT AAA CTC CAC ACA TCT ATC TTC G	181	
Glu Pro Thr Gly Thr Lys Leu His Thr Ser Ile Phe		
50                  55                  60		

## (2) INFORMATION FOR SEQ ID NO:140:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 60 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

Ser Gly Gln Gln Ser Leu Thr Gln Val Pro Gln Glu Ala Gly Arg Arg		
1                   5                   10                   15		
Ser Arg Thr His Val Thr His Gly Ile Pro His Ser Glu Ser Ser Gly		
20                  25                  30		
Thr Val Ser Arg Ser Asp Asp Pro Ser Ala Val Met Val Gln Val Ala		
35                  40                  45		

235

Glu Pro Thr Gly Thr Lys Leu His Thr Ser Ile Phe  
50 55 60

## (2) INFORMATION FOR SEQ ID NO:141:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 128 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: M62321 ORF1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

Met Gly Lys Val Pro Leu His Met Phe Leu Gln Val Leu Gly Pro Thr  
1 5 10 15

Ile Leu Ile Val Pro Phe Leu Thr Cys Pro Val Ile Ser Ala Pro Gln  
20 25 30

Trp Gln Arg Val Cys Met Met Pro Ser Thr Arg Gln Thr Pro Leu Tyr  
35 40 45

Pro Arg Trp Gln Asp Thr Lys Gly Ile Pro Gly Ser Cys Gly Met Ser  
50 55 60

Leu Ala Phe Ser Gln Val Leu Lys Ser Leu Asn Thr Ser His Ile Gln  
65 70 75 80

Ser Gln Met Ser Leu Ser Gln Glu Pro Glu His Gly Val Val His Ser  
85 90 95

Glu Leu Ile His Trp Cys Ser Arg Leu Arg Ser Trp Val Thr Val Arg  
100 105 110

236

Leu Leu Ser Met Ala Val Thr Arg Ala Ala Ala Ser Leu Ser Gly Thr  
115 120 125

## (2) INFORMATION FOR SEQ ID NO:142:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 144 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: M62321 ORF2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

Met Ala Gly Ser Arg Leu Thr Arg Ser Ser Val Glu Gly Thr Ser Pro  
1 5 10 15

Leu Met Ile Leu Asn Ala Thr Arg Ala Pro Ala Thr Pro Ala Pro Tyr  
20 25 30

Pro Ala Arg Met Ser Met Arg Thr Phe Pro Ser Pro Thr Leu Pro Met  
35 40 45

Ala Ala Pro Ala Lys Pro Ala Pro Thr Lys Ala Val Ala Ala Pro Gly  
50 55 60

Ala Ala Ser Trp Ala Ala Thr His Pro Pro Asn Met Leu Lys Arg Arg  
65 70 75 80

Val Trp Leu Val Val Ser Gly Leu Val Thr Ala Ala Val Lys Ala Ile  
85 90 95

Asn Glu Ala Met Ala Gly Leu Pro Gly Ser Val Asp Lys Pro Ala Lys  
100 105 110

237

Tyr Cys Ile Pro Leu Met Lys Phe His Ile Cys Phe Ala Gln Lys Val  
115 120 125

Ser Ser Phe Cys Gln Leu Val Trp Thr Ala Gly Ala Ile Thr Ser Ala  
130 135 140

(2) INFORMATION FOR SEQ ID NO:143:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 107 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: M58335, ORF1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

Met Gly Asn Val Pro Cys His Val Leu Leu Gln Val Leu Gly Pro Thr  
1 5 10 15

Ile Leu Met Glu Pro Phe Leu Thr Cys Pro Val Ile Cys Ala Pro His  
20 25 30

Gly Gln Val Val Cys Met Met Pro Ser Pro Arg Gln Thr Pro Leu Tyr  
35 40 45

Pro Arg Trp His Glu Lys Lys Gly Thr Pro Gly Ser Cys Gly Arg Ser  
50 55 60

Leu Asp Trp Ser Gln Val Leu Lys Ser Val Asn Thr Val His Ile Gln  
65 70 75 80

Ser Gln Thr Ser Leu Ser His Glu Pro Glu His Gly Val Glu Gln Ser  
85 90 95

238

Ser Leu Ile His Trp Trp Ser Leu Phe Ser Ser  
100 105

## (2) INFORMATION FOR SEQ ID NO:144:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 134 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: M58335, ORF2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:

Met Ser Thr Ser Thr Phe Pro Arg Pro Met Leu Pro Thr Ala Ala Pro  
1 5 10 15

Ala Met Pro Ala Pro Thr Lys Ala Glu Ala Ala Leu Gly Gly Ala Ser  
20 25 30

Trp Ala Ala Thr His Pro Pro Lys Met Leu Asn Arg Arg Val Leu Trp  
35 40 45

Val Val Ser Gly Leu Val Ile Glu Ala Val Asn Ala Ile Asn Asp Ala  
50 55 60

Ile Ala Gly Phe Pro Gly Arg Val Asp Lys Pro Ala Lys Tyr Cys Ile  
65 70 75 80

Pro Leu Met Lys Phe His Met Cys Phe Ala Gln Asn Val Ser Arg Ala  
85 90 95

Arg His Leu Asp Ser Thr Thr Gly Ala Ala Ala Ser Ala Cys Leu Val  
100 105 110

Ala Val Cys Ser Asn Pro Ser Ala Phe Cys Leu Asn Cys Ser Ala Ser

239

115

120

125

Cys Ile Pro Cys Ser Met  
130

## (2) INFORMATION FOR SEQ ID NO:145:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 100 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: D90208, ORF1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:

Met Gly Asn Val Pro Cys His Val Leu Leu Gln Val Phe Gly Pro Thr  
1 5 10 15

Ile Leu Met Glu Pro Phe Leu Thr Cys Pro Val Ile Cys Ala Pro His  
20 25 30

Gly Gln Val Val Cys Met Met Pro Ser Pro Arg Gln Thr Pro Leu Tyr  
35 40 45

Pro Arg Trp His Asp Arg Lys Gly Ser Pro Gly Asn Arg Gly Arg Ser  
50 55 60

Leu Asp Trp Ser Gln Val Leu Lys Ser Leu Asn Thr Val His Ile Gln  
65 70 75 80

Ser Gln Thr Ser Phe Ser His Glu Pro Glu Gln Gly Val Glu Gln Ser  
85 90 95

Ser Leu Ile His  
100

240

## (2) INFORMATION FOR SEQ ID NO:146:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 134 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: D90208, ORF2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

Met Ser Thr Ser Thr Phe Pro Arg Pro Met Leu Pro Thr Ala Ala Pro  
1 5 10 15

Ala Met Pro Ala Pro Thr Lys Ala Glu Ala Ala Leu Gly Gly Ala Ser  
20 25 30

Trp Ala Ala Thr His Pro Pro Lys Met Leu Asn Arg Arg Val Phe Trp  
35 40 45

Val Val Ser Gly Leu Val Ile Glu Ala Val Lys Ala Ile Asn Asp Ala  
50 55 60

Ile Ala Gly Phe Pro Gly Arg Val Asp Arg Pro Ala Lys Tyr Cys Ile  
65 70 75 80

Pro Leu Met Lys Phe His Met Cys Phe Ala Gln Lys Thr Ser Arg Ala  
85 90 95

Arg His Leu Asp Ser Thr Thr Gly Ala Ala Ala Ser Ala Cys Leu Val  
100 105 110

Ala Val Cys Ser Asn Pro Ser Ala Phe Cys Leu Asn Cys Ser Ala Ser  
115 120 125

Cys Ile Pro Cys Ser Met

## (2) INFORMATION FOR SEQ ID NO:147:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 175 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: Long Consensus Sequence, Fig. 11

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 16
- (D) OTHER INFORMATION: /note= "where X is G or S"

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 23
- (D) OTHER INFORMATION: /note= "where X is R or G"

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 38
- (D) OTHER INFORMATION: /note= "where X is C or R"

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 40
- (D) OTHER INFORMATION: /note= "where X is V or I"

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 44
- (D) OTHER INFORMATION: /note= "where X is P or S"

242

## (ix) FEATURE:

- (A) NAME/KEY: Modified-sit
- (B) LOCATION: 54
- (D) OTHER INFORMATION: /note= "where X is P or H"

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 65
- (D) OTHER INFORMATION: /note= "where X is M or T"

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 69
- (D) OTHER INFORMATION: /note= "where X is T or I"

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 92
- (D) OTHER INFORMATION: /note= "where X is T or A"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

His His His His Ser Ile Trp Pro Asp Val Arg Gly Gln Ala Pro Xaa  
1 5 10 15

Lys Gly Gln Gly Phe Gly Xaa Pro Pro Leu Pro Glu Gly Ala Pro His  
20 25 30

His Pro Leu Thr Asp Xaa Leu Xaa Pro Leu Thr Xaa Arg Pro Ile Asp  
35 40 45

His Gly Phe Val Pro Xaa Pro Pro Ala Leu Ile Glu Leu Arg Ser Ala  
50 55 60

Xaa Ala Gln Ala Xaa Thr Leu Ala Thr Thr Gln Pro Asn Thr Glu Val  
65 70 75 80

Ser Thr Ser Asn Val Ala Ser Lys Gln Asn Ser Xaa Pro Ser Thr Glu  
85 90 95

Val Arg Pro Arg Val Ala Glu Ile Pro Ile Lys Arg Ala Ser Gly Lys  
100 105 110

243

Ala Pro Arg Ala Ser Phe His Asn Thr Arg Asn Ile Arg Arg Trp Gly  
115 120 125

Pro Asn His Leu Lys Tyr Ser Thr Arg Phe Ala Lys Pro Asn Ile Ile  
130 135 140

Leu Thr Thr Gly Ser Pro Arg His Gln Asp Arg Ala Gly Pro Ala Glu  
145 150 155 160

Thr Ser Pro Ala Ala Ala Ser Thr Ala Gly Ser Pro Asn Leu  
165 170 175

(2) INFORMATION FOR SEQ ID NO:148:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Short Consensus Sequence, Fig. 11

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:148:

Thr Asn Arg Ala Ser Ala Ser  
1 5

(2) INFORMATION FOR SEQ ID NO:149:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

244

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Frame Shift Fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:

Pro	Pro	Ala	Leu	Ile	Glu	Leu	Arg	Ser	Ala	Met	Ala	Gln	Ala	Thr	Thr
1															
														10	
															15

(2) INFORMATION FOR SEQ ID NO:150:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 688 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: HGV Variant BG34

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 272..688

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:

GAATCGGCGC CGACTCGGCG ACCGGCCAAA AGGTGGTGG A TGGGTGATGA CAGGGTTGGT	60
AGGTCTAA TCCCCGTCAC CTTGGTAGCC ACTATAGGTG GGTCTTAAGA GAAAGCTTAAG	120
ATTCCCTCTTG TGCCTGCGGC GAGACCGCGC ACGGTCCACA GGTGTTGGCC CTACCGGTGT	180
GAATAAGGGC CCGACGTCAG GCTCGTCGTT AAACCGAGCC CGTCACCCAC CTGGGCCAAC	240
GACGCCACG TACGGTCCAC GTCGCCCTTC A ATG CCT CTC TTG GCC AAT AGG	292
Met Pro Leu Leu Ala Asn Arg	



246

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:

Met Pro Leu Leu Ala Asn Arg Ser Ile Arg Arg Val Asp Lys Asp Gln  
1 5 10 15

Trp Gly Pro Gly Val Thr Gly Met Asp Pro Gly Leu Cys Pro Ser Arg  
20 25 30

Trp Asn Gly Lys Arg Met Gly Pro Pro Ser Ser Ala Ala Ala Cys Ser  
35 40 45

Arg Gly Ser Pro Arg Thr Leu Arg Val Arg Ala Gly Gly Ile Ser Leu  
50 55 60

Phe Cys Ile Ile Met Ala Val Leu Leu Leu Leu Val Val Glu Ala  
65 70 75 80

Gly Ala Ile Leu Ala Pro Ala Thr His Ala Cys Arg Ala Asn Gly Gln  
85 90 95

Tyr Phe Leu Thr Asn Cys Cys Ala Leu Glu Asp Ile Gly Phe Cys Leu  
100 105 110

Glu Gly Gly Cys Leu Val Ala Leu Gly Cys Thr Ile Cys Thr Asp Arg  
115 120 125

Cys Trp Pro Leu Tyr Gln Ala Gly Leu Ala Val  
130 135

(2) INFORMATION FOR SEQ ID NO:152:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 663 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

247

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: HGV Variant T55806

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 271..663

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:

GACTCGGCGC CGACTCGCG ACCGGCCAAA AGGTGGTGG A	TGGGTGATGC CAGGGTTGGT	60	
AGGTCTGTAAG TCCCCTGTCAT CTTGGTAGCC ACTATACGTG GGTCTTAAGA GAAGGTTAAC		120	
ATTCCCTTTG TGCCCTGCCGC GAGACCGCGC ACGGTCCACA GGTGTTGGCC CTACCGGTGG		180	
AATAAGGGCC CGACGTCAGG CTCGTCGTTA AACCGAGCCC GTCACCCACC TGGGCCAAACG		240	
ACGCTCACGT ACGGTCCACG TCGCCCTTCA ATG TCT CTC TTG ACC AAT AGG TTT		294	
Met Ser Leu Leu Thr Asn Arg Phe			
1	5		
ATC CGG CGA GTT GAC AAG GAC CAG TGG GGG CCG GGG GTT ACG GGG ACG		342	
Ile Arg Arg Val Asp Lys Asp Gln Trp Gly Pro Gly Val Thr Gly Thr			
10	15	20	
GAC CCC GAA CCC TGC CCT TCC CCG TGG GCC GGG AAA TGC ATG GGG CCA		390.	
Asp Pro Glu Pro Cys Pro Ser Arg Trp Ala Gly Lys Cys Met Gly Pro			
25	30	35	40
CCC AGC TCC GCG GCG GCC TGC AGC CCG GGT AGC CCA AGA ATC CTT CGG		438	
Pro Ser Ser Ala Ala Cys Ser Arg Gly Ser Pro Arg Ile Leu Arg			
45	50	55	
GTG AGG GCG GGT GGC ATT TCT CTT TTC TAT ACC ATC ATG GCA GTC CTT		486	
Val Arg Ala Gly Gly Ile Ser Leu Phe Tyr Thr Ile Met Ala Val Leu			
60	65	70	
CTG CTC TTC TTC GTG GTT GAG GCC GGG GCG ATT CTC GCC CCG GCC ACC		534	
Leu Leu Phe Phe Val Val Glu Ala Gly Ala Ile Leu Ala Pro Ala Thr			
75	80	85	

248

CAC GCT TGT CGG GCG AAT GGG CAA TAT TTC CTC ACA AAT TGT TGC GCC	582		
His Ala Cys Arg Ala Asn Gly Gln Tyr Phe Leu Thr Asn Cys Cys Ala			
90	95	100	
CCA GAG GAT GTT GGG TTC TGC CTG GAG GGC GGA TGC CTG GTG GCT CTG	630		
Pro Glu Asp Val Gly Phe Cys Leu Glu Gly Gly Cys Leu Val Ala Leu			
105	110	115	120
GGG TGT ACG ATT TGC ACT GAC CGT TGC TGG CCA	663		
Gly Cys Thr Ile Cys Thr Asp Arg Cys Trp Pro			
125	130		

## (2) INFORMATION FOR SEQ ID NO:153:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 131 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:

Met Ser Leu Leu Thr Asn Arg Phe Ile Arg Arg Val Asp Lys Asp Gln			
1	5	10	15
Trp Gly Pro Gly Val Thr Gly Thr Asp Pro Glu Pro Cys Pro Ser Arg			
20	25	30	
Trp Ala Gly Lys Cys Met Gly Pro Pro Ser Ser Ala Ala Ala Cys Ser			
35	40	45	
Arg Gly Ser Pro Arg Ile Leu Arg Val Arg Ala Gly Gly Ile Ser Leu			
50	55	60	
Phe Tyr Thr Ile Met Ala Val Leu Leu Phe Phe Val Val Glu Ala			
65	70	75	80
Gly Ala Ile Leu Ala Pro Ala Thr His Ala Cys Arg Ala Asn Gly Gln			
85	90	95	
Tyr Phe Leu Thr Asn Cys Cys Ala Pro Glu Asp Val Gly Phe Cys Leu			
100	105	110	

249

Glu Gly Gly Cys Leu Val Ala Leu Gly Cys Thr Ile Cys Thr Asp Arg  
 115                    120                    125

Cys Trp Pro  
 130

## (2) INFORMATION FOR SEQ ID NO:154:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 632 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (C) INDIVIDUAL ISOLATE: HGV Variant EB20-2

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 271..632

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

GAATCGGCGC CGACTCGGCG ACCGGCCAAA AGGTGGTGGAA TGGGTGATGC CAGGGTTGGT	60
AGGTCTGTAAGA TCCCCTGTCAT CTTGGTAGCC ACTATAGGTG GGTCTTAAGA GAAGGTTAAG	120
ATTCCTCTTG TGCCTGCGGC GAGACCGCGC ACGGTCCACA GGTGTTGGCC CTACCGGTGT	180
AATAAGGGCC CGACGTCAGG CTCGTCGTTA AACCGAGGCC GTCACCCACC TGGGCAAACG	240
ACGCCAACGT ACGGTCCACG TCGCCCTTCA ATG CCT CTC TTG GCC AAT AGG AGT	294
Met Pro Leu Leu Ala Asn Arg Ser	
1                    5	
TAT CTC CGG CGA GTT GGC AAG GAC CAG TGG GGG CCG GGG GTT ACG GGG	342
Tyr Leu Arg Arg Val Gly Lys Asp Gln Trp Gly Pro Ile Val Thr Gly	

250

10	15	20	
AAG GAC CCC GAA CCC TGC CCT TCC CGG TGG GCC GGG AAA TGC ATG GGG Lys Asp Pro Glu Pro Cys Pro Ser Arg Trp Ala Gly Lys Cys Met Gly			390
25	30	35	40
CCA CCC AGC TCC GCG GCG GCC TGC AGC CGG GGT AGC CCA AAA AAC CTT Pro Pro Ser Ser Ala Ala Ala Cys Ser Arg Gly Ser Pro Lys Asn Leu			438
45	50	55	
CGG GTG AGG GCG GGT GGC ATT TTC TTT TCC TAT ACC ATC ATG GCA GTC Arg Val Arg Ala Gly Gly Ile Phe Phe Ser Tyr Thr Ile Met Ala Val			486
60	65	70	
CTT CTG CTC CTT CTC GTG GTT GAG GCC GGG GCC ATT TTG GCC CCG GCC Leu Leu Leu Leu Val Val Glu Ala Gly Ala Ile Leu Ala Pro Ala			534
75	80	85	
ACC CAC GCT TGC AGA GCT AAT GGG CAA TAT TTC CTC ACA AAC TGT TGT Thr His Ala Cys Arg Ala Asn Gly Gln Tyr Phe Leu Thr Asn Cys Cys			582
90	95	100	
GCC TTG GAG GAC ATC GGG TTC TGC CTG GAA GGC GGA TGC TTG GTG GCG CT Ala Leu Glu Asp Ile Gly Phe Cys Leu Glu Gly Gly Cys Leu Val Ala			632
105	110	115	120

## (2) INFORMATION FOR SEQ ID NO:155:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 120 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

Met Pro Leu Leu Ala Asn Arg Ser Tyr Leu Arg Arg Val Gly Lys Asp			
1	5	10	15
Gln Trp Gly Pro Gly Val Thr Gly Lys Asp Pro Glu Pro Cys Pro Ser			
20	25	30	

251

Arg Trp Ala Gly Lys Cys Met Gly Pro Pro Ser Ser Ala Ala Ala Cys  
35 40 45

Ser Arg Gly Ser Pro Lys Asn Leu Arg Val Val Arg Ala Gly Gly Ile Phe  
50 55 60

Phe Ser Tyr Thr Ile Met Ala Val Leu Leu Leu Leu Val Val Val Glu  
65 70 75 80

Ala Gly Ala Ile Leu Ala Pro Ala Thr His Ala Cys Arg Ala Asn Gly  
85 90 95

Gln Tyr Phe Leu Thr Asn Cys Cys Ala Leu Glu Asp Ile Gly Phe Cys  
100 105 110

Leu Glu Gly Gly Cys Leu Val Ala  
115 120

(2) INFORMATION FOR SEQ ID NO:156:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9103 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: HGV-JC Variant

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 276..9005

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

252

CAATGACTCG GCGCCGACTC GGCGACCGGC CAAAAGGTGG TGGATGGGTG ATGACAGGGT	60		
TGGTAGGTGCG TAAATCCCGG TCACCTTGGT AGCCACTATA GGTGGGTCTT AAGAGAAGGT	120		
TAAGATTCCCT CTTGTGCCTG CGGCGAGACC GCGCACGGTC CACAGGTGTT GGCCCTACCG	180		
GTGGGAATAA GGGCCCGACG TCAGGCTCGT CGTTAAACCG AGCCCCTAAC CGCCTGGGC	240		
AAACGACGCC CACGTACGGT CCACGTGCC CTTCA ATG TCG CTC TTG ACC AAT	293		
Met Ser Leu Leu Thr Asn			
1	5		
AGG CTT AGC CGG CGA GTT GAC AAG GAC CAG TGG GGG CCG GGG TTT ATG	341		
Arg Leu Ser Arg Arg Val Asp Lys Asp Gln Trp Gly Pro Gly Phe Met			
10	15	20	
GGG AAG GAC CCC AAA CCC TGC CCT TCC CGG CGG ACC GGG AAA TGC ATG	389		
Gly Lys Asp Pro Lys Pro Cys Pro Ser Arg Arg Thr Gly Lys Cys Met			
25	30	35	
GGG CCA CCC AGC TCC GCG GCG GCC TGC AGC CGG GGT AGC CCA AGA ATC	437		
Gly Pro Pro Ser Ser Ala Ala Ala Cys Ser Arg Gly Ser Pro Arg Ile			
40	45	50	
CTT CGG GTG AGG GCG GGT GGC ATT TCT CTT CCT TAT ACC ATC ATG GAA	485		
Leu Arg Val Arg Ala Gly Gly Ile Ser Leu Pro Tyr Thr Ile Met Glu			
55	60	65	70
GCC CTC CTG TTC CTC CTC GGG GTG GAG GGC GGG GCC ATT CTG GCC CCG	533		
Ala Leu Leu Phe Leu Leu Gly Val Glu Ala Gly Ala Ile Leu Ala Pro			
75	80	85	
GCC ACC CAC GCT TGT CGA GCG AAT GGG CAA TAT TTC CTC ACA AAC TGT	581		
Ala Thr His Ala Cys Arg Ala Asn Gly Gln Tyr Phe Leu Thr Asn Cys			
90	95	100	
TGT GCT CCA GAG GAC ATT GGG TTC TGC CTC GAA GGC GGT TGC CTT GTG	629		
Cys Ala Pro Glu Asp Ile Gly Phe Cys Leu Glu Gly Gly Cys Leu Val			
105	110	115	
GCC CTG GGG TGC ACA GTT TGC ACT GAC CGA TGC TGG CCG CTG TAT CAG	677		
Ala Leu Gly Cys Thr Val Cys Thr Asp Arg Cys Trp Pro Leu Tyr Gln			
120	125	130	

253

GCG GGC TTG GCT GTG CGG CCT GGC AAG TCC GCA GCC CAG CTG GTG GGG Ala Gly Leu Ala Val Arg Pro Gly Lys Ser Ala Ala Gln Leu Val Gly 135                  140                  145                  150	725
CAA CTG GGT GGC CTC TAC GGG CCC TTG TCG GTG TCG GCC TAC GTG GCC Gln Leu Gly Gly Leu Tyr Gly Pro Leu Ser Val Ser Ala Tyr Val Ala 155                  160                  165	773
GGC ATC CTG GGC CTG GGT GAG GTG TAC TCG GGT GTC CTA ACA GTT GGT Gly Ile Leu Gly Leu Gly Glu Val Tyr Ser Gly Val Leu Thr Val Gly 170                  175                  180	821
GTT GCG TTG ACG CGC CCG GTC TAC CCG ATG CCC AAC CTG ACG TGT GCA Val Ala Leu Thr Arg Arg Val Tyr Pro Met Pro Asn Leu Thr Cys Ala 185                  190                  195	869
GTA GAG TGT GAG CTT AAG TGG GAA AGT GAG TTT TGG AGA TGG ACT GAG Val Glu Cys Glu Leu Lys Trp Glu Ser Glu Phe Trp Arg Trp Thr Glu 200                  205                  210	917
CAG CTG GCC TCC AAT TAC TGG ATT CTG GAA TAC CTT TGG AAG GTC CCG Gln Leu Ala Ser Asn Tyr Trp Ile Leu Glu Tyr Leu Trp Lys Val Pro 215                  220                  225                  230	965
TTT GAC TTC TGG AGA GGC GTG CTA AGC CTG ACT CCC TTG CTG GTT TGC Phe Asp Phe Trp Arg Gly Val Leu Ser Leu Thr Pro Leu Leu Val Cys 235                  240                  245	1013
GTG GCC GCG TTG CTG CTG CTG GAG CAA CGG ATT GTC ATG GTC TTC CTG Val Ala Ala Leu Leu Leu Glu Gln Arg Ile Val Met Val Phe Leu 250                  255                  260	1061
TTG GTG ACG ATG GCC GGG ATG TCG CAA GGC GCT CCG GCC TCC GTT TTG Leu Val Thr Met Ala Gly Met Ser Gln Gly Ala Pro Ala Ser Val Leu 265                  270                  275	1109
GGG TCT CGC CCC TTT GAC TAC GGG TTG ACA TGG CAG TCT TGT TCC TGC Gly Ser Arg Pro Phe Asp Tyr Gly Leu Thr Trp Gln Ser Cys Ser Cys 280                  285                  290	1157
AGG GCT AAT GGG TCG CGC TAT ACT ACT GGG GAG AAG GTG TGG GAC CGT Arg Ala Asn Gly Ser Arg Tyr Thr Thr Gly Glu Lys Val Trp Asp Arg 295                  300                  305                  310	1205

254

GGG AAC GTC ACG CTC CTG TGT GAC TGC CCC AAC GGC CCC TGG GTG TGG Gly Asn Val Thr Leu Leu Cys Asp Cys Pro Asn Gly Pro Trp Val Trp	315	320	325	1253	
TTG CCG GCC TTT TGC CAA GCA ATC GGC TGG GGC GAT CCC ATC ACT CAT Leu Pro Ala Phe Cys Gln Ala Ile Gly Trp Gly Asp Pro Ile Thr His	330	335	340	1301	
TGG AGC CAC GGC CAA AAT CGG TGG CCC CTC TCA TGC CCC CAG TAT GTC Trp Ser His Gly Gln Asn Arg Trp Pro Leu Ser Cys Pro Gln Tyr Val	345	350	355	1349	
TAT GGG TCT GTT TCA GTC ACT TGC GTG TGG GGT TCC GTC TCT TGG TTT Tyr Gly Ser Val Ser Val Thr Cys Val Trp Gly Ser Val Ser Trp Phe	360	365	370	1397	
GCC TCG ACT GGC GGT CGC GAC TCG AAG ATC GAT GTG TGG AGT CTG GTG Ala Ser Thr Gly Gly Arg Asp Ser Lys Ile Asp Val Trp Ser Leu Val	375	380	385	390	1445
CCG GTT GGT TCC GCC AGC TGC ACC ATA GCC GCT CTT GGA TCG TCG GAT Pro Val Gly Ser Ala Ser Cys Thr Ile Ala Ala Leu Gly Ser Ser Asp	395	400	405	1493	
CGG GAC ACG GTA GTT GAG CTC TCC GAG TGG GGA GTC CCG TGC GCA ACG Arg Asp Thr Val Val Glu Leu Ser Glu Trp Gly Val Pro Cys Ala Thr	410	415	420	1541	
TGC ATT CTG GAT CGT CCG CCG GCC TCG TGC GGC ACC TGT GTG AGA GAC Cys Ile Leu Asp Arg Arg Pro Ala Ser Cys Gly Thr Cys Val Arg Asp	425	430	435	1589	
TGC TGG CCC GAA ACC GGG TCG GTT AGG TTT CCA TTC CAT CGG TGC GGC Cys Trp Pro Glu Thr Gly Ser Val Arg Phe Pro Phe His Arg Cys Gly	440	445	450	1637	
GCG GGG CCT AAG CTG ACA AAG GAC TTG GAA GCT GTG CCC TTC GTC AAT Ala Gly Pro Lys Leu Thr Lys Asp Leu Glu Ala Val Pro Phe Val Asn	455	460	465	470	1685
AGG ACA ACT CCC TTC ACC ATA AGG GGC CCC CTG GGC AAC CAG GGG AGA Arg Thr Thr Pro Phe Thr Ile Arg Gly Pro Leu Gly Asn Gln Gly Arg	475	480	485	1733	

255

GGC AAC CCG GTG CGG TCG CCC TTG GGT TTT GGG TCC TAC GCC ATG ACC      1781  
 Gly Asn Pro Val Arg Ser Pro Leu Gly Phe Gly Ser Tyr Ala Met Thr  
 490                  495                  500

AAG ATC CGA GAC TCC TTA CAT TTG GTG AAA TGT CCC ACA CCA GCC ATT      1829  
 Lys Ile Arg Asp Ser Leu His Leu Val Lys Cys Pro Thr Pro Ala Ile  
 505                  510                  515

GAG CCT CCC ACC GGG ACG TTT GGG TTC TTC CCC GGA GTG CCG CCT CTT      1877  
 Glu Pro Pro Thr Gly Thr Phe Gly Phe Phe Pro Gly Val Pro Pro Leu  
 520                  525                  530

AAC AAC TGC CTG CTG TTG GGC ACG GAA GTG TCC GAA GCG CTG GGC GGG      1925  
 Asn Asn Cys Leu Leu Gly Thr Glu Val Ser Glu Ala Leu Gly Gly  
 535                  540                  545                  550

GCC GGC CTC ACG GGG GGG TTC TAT GAA CCC CTG GTG CGC AGG CGT TCG      1973  
 Ala Gly Leu Thr Gly Gly Phe Tyr Glu Pro Leu Val Arg Arg Arg Ser  
 555                  560                  565

GAG CTG ATG GGG CGC CGA AAT CCG GTT TGC CCG GGG TTT GCA TGG CTG      2021  
 Glu Leu Met Gly Arg Arg Asn Pro Val Cys Pro Gly Phe Ala Trp Leu  
 570                  575                  580

TCC TCG GGT CGA CCT GAC GGG TTT ATA CAC GTC CAG GGC CAC TTG CAG      2069  
 Ser Ser Gly Arg Pro Asp Gly Phe Ile His Val Gln Gly His Leu Gln  
 585                  590                  595

GAG GTC GAT GCT GGC AAC TTC ATC CCT CCA CCT CGC TGG TTG CTC TTG      2117  
 Glu Val Asp Ala Gly Asn Phe Ile Pro Pro Arg Trp Leu Leu Leu  
 600                  605                  610

GAC TTT GTG TTT GTC CTG TTA TAC CTG ATG AAG CTG GCT GAG GCA CGG      2165  
 Asp Phe Val Phe Val Leu Leu Tyr Leu Met Lys Leu Ala Glu Ala Arg  
 615                  620                  625                  630

CTG GTC CCG TTG ATC TTG CTT CTG CTG TGG TGG TGG GTG AAC CAG TTG      2213  
 Leu Val Pro Leu Ile Leu Leu Leu Trp Trp Trp Val Asn Gln Leu  
 635                  640                  645

GCA GTC CTT GGA CTG CCG GCT GTG GAC GCC GCC GTG GCT GGT GAG GTC      2261  
 Ala Val Leu Gly Leu Pro Ala Val Asp Ala Ala Val Ala Gly Glu Val  
 650                  655                  660

256

TTC GCG GGC CCG GCC CTG TCG TGG TGT CTG GGC CTC CCC ACC GTT AGT Phe Ala Gly Pro Ala Leu Ser Trp Cys Leu Gly Leu Pro Thr Val Ser 665                   670                   675	2309
ATG ATC CTG GGC TTA GCA AAC CTG GTG TTG TAT TTC CGG TGG ATG GGT Met Ile Leu Gly Leu Ala Asn Leu Val Leu Tyr Phe Arg Trp Met Gly 680                   685                   690	2357
CCC CAA CGC CTC ATG TTC CTC GTG TTG TGG AAG CTC GCT CGG GGA GCC Pro Gln Arg Leu Met Phe Leu Val Leu Trp Lys Leu Ala Arg Gly Ala 695                   700                   705                   710	2405
TTC CCG CTG GCA CTT CTG ATG GGG ATC TCG GCA ACC CGC GGG CGC ACC Phe Pro Leu Ala Leu Leu Met Gly Ile Ser Ala Thr Arg Gly Arg Thr 715                   720                   725	2453
TCG GTG CTC GGG GCC GAG TTC TGC TTC GAT GTC ACA TTC GAG GTG GAC Ser Val Leu Gly Ala Glu Phe Cys Phe Asp Val Thr Phe Glu Val Asp 730                   735                   740	2501
ACG TCG GTT TTG GGC TGG GTG GTG GCC AGT GTG GTA GCC TGG GCC ATT Thr Ser Val Leu Gly Trp Val Val Ala Ser Val Val Ala Trp Ala Ile 745                   750                   755	2549
GCG CTC CTG AGC TCG ATG AGC GCG GGA GGG TGG AGG CAC AAG GCC GTG Ala Leu Leu Ser Ser Met Ser Ala Gly Gly Trp Arg His Lys Ala Val 760                   765                   770	2597
ATC TAT AGG ACG TGG TGT AAG GGG TAC CAG GCA ATA CGC CAA CGG GTG Ile Tyr Arg Thr Trp Cys Lys Gly Tyr Gln Ala Ile Arg Gln Arg Val 775                   780                   785                   790	2645
GTG CGG AGC CCC CTC GGG GAG GGG CGG CCC ACC AAA CCC TTG ACG TTT Val Arg Ser Pro Leu Gly Glu Gly Arg Pro Thr Lys Pro Leu Thr Phe 795                   800                   805	2693
GCT TGG TGC TTG GCC TCA TAC ATC TGG CCG GAT GCT GTG ATG ATG GTG Ala Trp Cys Leu Ala Ser Tyr Ile Trp Pro Asp Ala Val Met Met Val 810                   815                   820	2741
GTG GTA GCC TTG GTG CTC CTC TTT GGC CTG TTC GAC GCG TTG GAC TGG Val Val Ala Leu Val Leu Leu Phe Gly Leu Phe Asp Ala Leu Asp Trp 825                   830                   835	2789

257

GCT TTG GAG GAG CTC TTG GTG TCC CGG CCC TCG TTA CGG CGT CTG GCC Ala Leu Glu Glu Leu Leu Val Ser Arg Pro Ser Leu Arg Arg Leu Ala 840 845 850	2837
CGG GTG GTT GAG TGC TGT GTG ATG GCG GGA GAG AAG GCC ACA ACC GTC Arg Val Val Glu Cys Cys Val Met Ala Gly Glu Lys Ala Thr Thr Val 855 860 865 870	2885
CGG CTG GTC TCC AAG ATG TGC GCG AGA GGG GCC TAT TTG TTT GAC CAT Arg Leu Val Ser Lys Met Cys Ala Arg Gly Ala Tyr Leu Phe Asp His 875 880 885	2933
ATG GGC TCT TTT TCG CGC GCT GTC AAG GAG CGC CTG CTG GAG TGG GAC Met Gly Ser Phe Ser Arg Ala Val Lys Glu Arg Leu Leu Glu Trp Asp 890 895 900	2981
GCG GCT TTG GAA CCC CTG TCA TTC ACT AGG ACG GAC TGT CGC ATC ATT Ala Ala Leu Glu Pro Leu Ser Phe Thr Arg Thr Asp Cys Arg Ile Ile 905 910 915	3029
AGA GAT GCT GCG AGG ACC TTG GCC TGC GGG CAG TGC GTC ATG GGC TTG Arg Asp Ala Ala Arg Thr Leu Ala Cys Gly Gln Cys Val Met Gly Leu 920 925 930	3077
CCT GTG GTA GCG CGC CGT GGT GAC GAG GTT CTT ATC GGT GTC TTT CAG Pro Val Val Ala Arg Arg Gly Asp Glu Val Leu Ile Gly Val Phe Gln 935 940 945 950	3125
GAT GTG AAC CAT TTG CCT CCC CGA TTC GTC CCG ACC GCA CCC GTT GTC Asp Val Asn His Leu Pro Pro Gly Phe Val Pro Thr Ala Pro Val Val 955 960 965	3173
ATC CGG CGG TGC GGG AAG GGG TTT CTG GGG GTC ACT AAG GCT GCC TTG Ile Arg Arg Cys Gly Lys Gly Phe Leu Gly Val Thr Lys Ala Ala Leu 970 975 980	3221
ACT GGT CGG GAT CCT GAC TTA CAT CCA GGG AAC GTC ATG GTG TTG GGG Thr Gly Arg Asp Pro Asp Leu His Pro Gly Asn Val Met Val Leu Gly 985 990 995	3269
ACG GCT ACG TCG CGA AGC ATG GGG ACA TGC CTG AAC GGC CTG CTG TTC Thr Ala Thr Ser Arg Ser Met Gly Thr Cys Leu Asn Gly Leu Leu Phe 1000 1005 1010	3317

258

ACG ACT TTC CAT GGG GCT TCA TCC CGA ACC ATC GCC ACG CCC GTG GGG Thr Thr Phe His Gly Ala Ser Ser Arg Thr Ile Ala Thr Pro Val Gly 1015                1020                1025                1030	3365
GCC CTT AAT CCC AGG TGG TGG TCC GCC AGT GAT GAC GTC ACG GTG TAC Ala Leu Asn Pro Arg Trp Trp Ser Ala Ser Asp Asp Val Thr Val Tyr 1035                1040                1045	3413
CCG CTC CCG GAT GGG GCA ACC TCG TTG ACG CCC TGC ACT TGC CAG GCT Pro Leu Pro Asp Gly Ala Thr Ser Leu Thr Pro Cys Thr Cys Gln Ala 1050                1055                1060	3461
GAG TCC TGT TGG GTC ATA CGG TCC GAC GGG GCT TTG TGC CAT GCC TTG Glu Ser Cys Trp Val Ile Arg Ser Asp Gly Ala Leu Cys His Gly Leu 1065                1070                1075	3509
AGT AAG GGA GAC AAG GTG GAG CTA GAT GTG GCC ATG GAG GTC TCA GAT Ser Lys Gly Asp Lys Val Glu Leu Asp Val Ala Met Glu Val Ser Asp 1080                1085                1090	3557
TTC CGT GGC TCG TCC GGC TCA CCT GTC CTG TGC GAC GAG GGG CAC GCA Phe Arg Gly Ser Ser Gly Ser Pro Val Leu Cys Asp Glu Gly His Ala 1095                1100                1105                1110	3605
GTA GGA ATG CTC GTG TCG CTC CAC TCG GGT CGG GTC ACC GCG Val Gly Met Leu Val Ser Val Leu His Ser Gly Gly Arg Val Thr Ala 1115                1120                1125	3653
GCT CGA TTC ACC AGG CCG TGG ACC CAG GTC CCA ACA GAT GCT AAG ACC Ala Arg Phe Thr Arg Pro Trp Thr Gln Val Pro Thr Asp Ala Lys Thr 1130                1135                1140	3701
ACC ACT GAA CCC CCT CCG GTG CCG GCA AAG GGA GTT TTC AAG GAA GCC Thr Thr Glu Pro Pro Val Pro Ala Lys Gly Val Phe Lys Glu Ala 1145                1150                1155	3749
CCA CTG TTT ATG CCC ACG GGC GCA GGA AAG AGC ACG CGC GTC CCG TTG Pro Leu Phe Met Pro Thr Gly Ala Gly Lys Ser Thr Arg Val Pro Leu 1160                1165                1170	3797
GAG TAT GGC AAC ATG GGG CAC AAG GTC CTG ATT TTG AAC CCC TCG GTG Glu Tyr Gly Asn Met Gly His Lys Val Leu Ile Leu Asn Pro Ser Val 1175                1180                1185                1190	3845

259

GCG ACA GTG AGG GCC ATG GGC CCT TAC ATG GAG CGA CTG GCG GGA AAA Ala Thr Val Arg Ala Met Gly Pro Tyr Met Glu Arg Leu Ala Gly Lys 1195	1200	1205	3893	
CAT CCA AGT ATC TAC TGT GGC CAT GAC ACC ACT GCC TTC ACA AGG ATC His Pro Ser Ile Tyr Cys Gly His Asp Thr Thr Ala Phe Thr Arg Ile 1210	1215	1220	3941	
ACT GAT TCC CCC TTA ACG TAC TCT ACC TAT GGG AGG TTT CTG GCC ACG Thr Asp Ser Pro Leu Thr Tyr Ser Thr Tyr Gly Arg Phe Leu Ala Asn 1225	1230	1235	3989	
CCT AGG CAG ATG CTG CGA GGT GTG TCG GTG GTC ATT TGC GAT GAA TCC Pro Arg Gln Met Leu Arg Gly Val Ser Val Val Ile Cys Asp Glu Cys 1240	1245	1250	4037	
CAC AGT CAT GAT TCC ACT GTG TTG TTG GGG ATT GGA CGG GTC CGG GAC His Ser His Asp Ser Thr Val Leu Leu Gly Ile Gly Arg Val Arg Glu 1255	1260	1265	1270	4085
CTG GCA CGA GAG TGT GGG GTG CAG CTT GTG CTC TAC GCC ACT GCC ACG Leu Ala Arg Glu Cys Gly Val Gln Leu Val Leu Tyr Ala Thr Ala Thr 1275	1280	1285	4133	
CCT CCT GGG TCC CCC ATG ACT CAG CAT CCG TCA ATC ATT GAG ACC AAA Pro Pro Gly Ser Pro Met Thr Gln His Pro Ser Ile Ile Glu Thr Lys 1290	1295	1300	4181	
TTG GAT GTG GGT GAG ATT CCC TTC TAT GGG CAT GGC ATA CCC CTC GAG Leu Asp Val Gly Glu Ile Pro Phe Tyr Gly His Gly Ile Pro Leu Glu 1305	1310	1315	4229	
CGG ATG CGG ACC GGT AGG CAC CTC GTA TTC TGC TAC TCT AAG GCA GAG Arg Met Arg Thr Gly Arg His Leu Val Phe Cys Tyr Ser Lys Ala Glu 1320	1325	1330	4277	
TGT GAG CGG CTA GCC GGT CAG TTT TCT GCT AGG GGA GTT AAC GCC ATA Cys Glu Arg Leu Ala Gly Gln Phe Ser Ala Arg Gly Val Asn Ala Ile 1335	1340	1345	1350	4325
GCC TAT TAC AGG GGA AAA GAC AGT TCT ATC ATC AAG GAC GGA GAT CTG Ala Tyr Tyr Arg Gly Lys Asp Ser Ser Ile Ile Lys Asp Gly Asp Leu 1355	1360	1365	4373	

260

GTG GTG TGC GCG ACC GAC GCG CTA TCC ACT GGA TAC ACT GGG AAC TTC Val Val Cys Ala Thr Asp Ala Leu Ser Thr Gly Tyr Thr Gly Asn Phe 1370 1375 1380	4421
GAT TCT GTC ACC GAC TGT GGG TTA GTG GTG GAG GAG GTC GTC GAG GTG Asp Ser Val Thr Asp Cys Gly Leu Val Val Glu Glu Val Val Glu Val 1385 1390 1395	4469
ACC CTT GAT CCC ACC ATT ACC ATC TCC CTG CCG ACA GTG CCC GCG TCG Thr Leu Asp Pro Thr Ile Thr Ile Ser Leu Arg Thr Val Pro Ala Ser 1400 1405 1410	4517
GCA GAA CTG TCG ATG CAG AGA CGA GGA CGC ACG GGT AGA GGC AGG TCT Ala Glu Leu Ser Met Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Ser 1415 1420 1425 1430	4565
GGG CGC TAC TAC TAC GCC GGG GTC GGA AAG GCC CCC GCG GGT GTG GTG Gly Arg Tyr Tyr Tyr Ala Gly Val Gly Lys Ala Pro Ala Gly Val Val 1435 1440 1445	4613
CGC TCG GGT CCT GTC TGG TCG GCG GTG GAG GCC GGA GTG ACC TGG TAT Arg Ser Gly Pro Val Trp Ser Ala Val Glu Ala Gly Val Thr Trp Tyr 1450 1455 1460	4661
GGA ATG GAA CCT GAC TTG ACA GCT AAC CTA TTG AGA CTT TAC GAC GAC Gly Met Glu Pro Asp Leu Thr Ala Asn Leu Leu Arg Leu Tyr Asp Asp 1465 1470 1475	4709
TGC CCT TAC ACC GCA GCC GTC GCA GCT GAC ATC GGT GAA GCC GCG GTG Cys Pro Tyr Thr Ala Ala Val Ala Asp Ile Gly Glu Ala Ala Val 1480 1485 1490	4757
TTT TTC TCC CGG CTA GCC CCG TTG AGC ATG CAT CCC GAT GTT AGC TGG Phe Phe Ser Gly Leu Ala Pro Leu Arg Met His Pro Asp Val Ser Trp 1495 1500 1505 1510	4805
GCA AAA GTG CGC GGC GTC AAC TGG CCC CTC TTG GTG GGT GTT CAG CGG Ala Lys Val Arg Gly Val Asn Trp Pro Leu Leu Val Gly Val Gln Arg 1515 1520 1525	4853
ACC ATG TGC CGG GAA ACA CTG TCT CCC GGA CCA TCG GAC GAC CCC CAA Thr Met Cys Arg Glu Thr Leu Ser Pro Gly Pro Ser Asp Asp Pro Gln 1530 1535 1540	4901

261

TGG GCA GGT CTG AAG GGC CCG AAT CCT GTT CCA CTA CTG CTG AGG TGG Trp Ala Gly Leu Lys Gly Pro Asn Pro Val Pro Leu Leu Arg Trp	4949
1545 1550 1555	
GGC AAT GAT TTA CCA TCA AAA GTG GCC GGC CAC CAC ATT GTT GAC GAC Gly Asn Asp Leu Pro Ser Lys Val Ala Gly His His Ile Val Asp Asp	4997
1560 1565 1570	
CTG GTT CGT AGG CTT GGT GTG GCG GAG GGT TAT GTC CGC TGC GAT GCG Leu Val Arg Arg Leu Gly Val Ala Glu Gly Tyr Val Arg Cys Asp Ala	5045
1575 1580 1585 1590	
GGG CCG ATC TTA ATG GTC GGC CTC GCT ATC GCG GGG GGG ATG ATC TAC Gly Pro Ile Leu Met Val Gly Leu Ala Ile Ala Gly Gly Met Ile Tyr	5093
1595 1600 1605	
GCA TCT TAC ACC GGG TCT TTA GTG GTG GTG ACA GAC TGG GAT GTA AAG Ala Ser Tyr Thr Gly Ser Leu Val Val Val Thr Asp Trp Asp Val Lys	5141
1610 1615 1620	
GGG GGT GGC AGC CCT CTT TAT CCG CAT GGA GAC CAG GCC ACG CCA CAG Gly Gly Gly Ser Pro Leu Tyr Arg His Gly Asp Gln Ala Thr Pro Gln	5189
1625 1630 1635	
CCG GTT GTG CAG GTC CCC CCG GTA GAC CAT CGG CCG GGG GGG GAG TCT Pro Val Val Gln Val Pro Pro Val Asp His Arg Pro Gly Glu Ser	5237
1640 1645 1650	
GCG CCT TCG GAT GCC AAG ACA GTG ACA GAT GCG GTG GCG GCC ATC CAG Ala Pro Ser Asp Ala Lys Thr Val Thr Asp Ala Val Ala Ala Ile Gln	5285
1655 1660 1665 1670	
GTG GAT TGC GAT TGG TCA GTC ATG ACC CTG TCG ATC GGG GAA GTG CTG Val Asp Cys Asp Trp Ser Val Met Thr Leu Ser Ile Gly Glu Val Leu	5333
1675 1680 1685	
TCC TTG GCT CAG GCT AAA ACA GCT GAG GCC TAC ACG GCA ACC GCC AAG Ser Leu Ala Gln Ala Lys Thr Ala Glu Ala Tyr Thr Ala Thr Ala Lys	5381
1690 1695 1700	
TGG CTC GCT GGC TGC TAC ACG GGG ACG CGG GCC GTT CCC ACT GTT TCA Trp Leu Ala Gly Cys Tyr Thr Gly Thr Arg Ala Val Pro Thr Val Ser	5429
1705 1710 1715	

262

ATT GTT GAC AAG CTC TTT GCC GGA GGG TGG GCG GCT GTG GTT GGC CAC Ile Val Asp Lys Leu Phe Ala Gly Gly Trp Ala Ala Val Val Gly His 1720                    1725                    1730	5477
TGT CAC AGC GTC ATA GCT GCG GCG GTG GCT GCC TAC GGG GCT TCC AGG Cys His Ser Val Ile Ala Ala Val Ala Ala Tyr Gly Ala Ser Arg 1735                    1740                    1745                    1750	5525
AGT CCG CCG TTG GCA GCC GCG GCT TCC TAC CTG ATG GGA CTG GGC GTC Ser Pro Pro Leu Ala Ala Ala Ser Tyr Leu Met Gly Leu Gly Val 1755                    1760                    1765	5573
GGA GGC AAC GCT CAG ACG CGT TTG GCG TCT GCC CTC CTG TTG GGG GCC Gly Gly Asn Ala Gln Thr Arg Leu Ala Ser Ala Leu Leu Leu Gly Ala 1770                    1775                    1780	5621
GCT GGC ACC GCC CTG GGC ACT CCC GTC GTG GGT TTA ACC ATG GCG GGG Ala Gly Thr Ala Leu Gly Thr Pro Val Val Gly Leu Thr Met Ala Gly 1785                    1790                    1795	5669
GCG TTC ATG GGG GGT GCT AGC GTC TCT CCC TCC TTG GTC ACC ATC ATC TTG Ala Phe Met Gly Ala Ser Val Ser Pro Ser Leu Val Thr Ile Leu 1800                    1805                    1810	5717
TG GGG GCC GTG GGA GGC TGG GAG GGC GTC GTC AAC GCT GCT AGC CTT Leu Gly Ala Val Gly Gly Trp Glu Gly Val Val Asn Ala Ala Ser Leu 1815                    1820                    1825                    1830	5765
GTC TTT GAC TTC ATG GCG GGG AAA CTA TCG TCA GAA GAT CTG TGG TAC Val Phe Asp Phe Met Ala Gly Lys Leu Ser Ser Glu Asp Leu Trp Tyr 1835                    1840                    1845	5813
GCC ATC CCA GTG CTC ACC AGC CCG GGG GCG GGC CTT GCG GGG ATC GCC Ala Ile Pro Val Leu Thr Ser Pro Gly Ala Gly Leu Ala Gly Ile Ala 1850                    1855                    1860	5861
CTT GGG TTG GTG CTG TAC TCA GCT AAC AAC TCT GGT ACT ACC ACT TGG Leu Gly Leu Val Leu Tyr Ser Ala Asn Asn Ser Gly Thr Thr Thr Trp 1865                    1870                    1875	5909
TTG AAC CGT CTG CTG ACT ACG TTA CCT AGG TCT TCT TGC ATC CCT GAC Leu Asn Arg Leu Leu Thr Thr Leu Pro Arg Ser Ser Cys Ile Pro Asp 1880                    1885                    1890	5957

## 263

AGC TAT TTC CAA CAG GCC GAT TAC TGT GAC AAG GTC TCG GCC GTG CTT Ser Tyr Phe Gln Gln Ala Asp Tyr Cys Asp Lys Val Ser Ala Val Leu	1895                  1900                  1905                  1910	6005
CGC CGA CTG AGC CTC ACC CGC ACT GTG GTG GCC CTA GTC AAT AGG GAA Arg Arg Leu Ser Leu Thr Arg Thr Val Val Ala Leu Val Asn Arg Glu	1915                  1920                  1925	6053
CCC AAG GTG GAC GAG GTA CAG GTG GGG TAC GTC TGG GAT CTC TGG GAG Pro Lys Val Asp Glu Val Gln Val Gly Tyr Val Trp Asp Leu Trp Glu	1930                  1935                  1940	6101
TGG ATC ATG CGT CAA GTG CGC ATG GTC ATG GCC AGG CTC CCG CCT CTC Trp Ile Met Arg Gln Val Arg Met Val Met Ala Arg Leu Arg Ala Leu	1945                  1950                  1955	6149
TGC CCC GTG GTG TCA CTG CCT TTG TGG CAC TGC GGG GAG GGG TGG TCC Cys Pro Val Val Ser Leu Pro Leu Trp His Cys Gly Glu Gly Trp Ser	1960                  1965                  1970	6197
GGA GAG TGG TTG TTG GAC GGC CAT GTG GAG AGT CGC TGT CTT TGC GGG Gly Glu Trp Leu Leu Asp Gly His Val Glu Ser Arg Cys Leu Cys Gly	1975                  1980                  1985                  1990	6245
TGC GTG ATC ACC GGC GAT GTT TTC AAT GGG CAA CTC AAA GAG CCA GTT Cys Val Ile Thr Gly Asp Val Phe Asn Gly Gln Leu Lys Glu Pro Val	1995                  2000                  2005	6293
TAC TCT ACA AAG TTG TGC CGG CAC TAT TGG ATG GGG ACC GTT CCT GTG Tyr Ser Thr Lys Leu Cys Arg His Tyr Trp Met Gly Thr Val Pro Val	2010                  2015                  2020	6341
AAC ATG CTG GGT TAC GGC GAA ACA TCA CCC CTC TTG GCC TCT GAC ACC Asn Met Leu Gly Tyr Gly Glu Thr Ser Pro Leu Leu Ala Ser Asp Thr	2025                  2030                  2035	6389
CCG AAG GTG GTG CCT TTT GGG ACG TCG GGC TGG GCT GAG GTG GTG GTG Pro Lys Val Val Pro Phe Gly Thr Ser Gly Trp Ala Glu Val Val Val	2040                  2045                  2050	6437
ACC CCT ACC CAC GTG GTG ATC AGG AGA ACC TCT CCC TAC GAG TTG CTG Thr Pro Thr His Val Val Ile Arg Arg Thr Ser Pro Tyr Glu Leu Leu	2055                  2060                  2065                  2070	6485

264

CGC CAA CAA ATC CTA TCA GCT GCA GTT GCT GAG CCC TAT TAT GTC GAC        6533  
 Arg Gln Gln Ile Leu Ser Ala Ala Val Ala Glu Pro Tyr Tyr Val Asp  
 2075                  2080                  2085

GGC ATA CCG GTC TCA TGG GAC GCG GAC GCT CGT GCG CCT GCT ATG GTT        6581  
 Gly Ile Pro Val Ser Trp Asp Ala Asp Ala Arg Ala Pro Ala Met Val  
 2090                  2095                  2100

TAT GGC CCT GGG CAA AGT GTT ACC ATT GAC GGG GAG CGC TAC ACC CTC        6629  
 Tyr Gly Pro Gly Gln Ser Val Thr Ile Asp Gly Glu Arg Tyr Thr Leu  
 2105                  2110                  2115

CCG CAT CAA CTG CGG CTC AGG AAT GTA GCG CCC TCT GAG GTT TCA TCC        6677  
 Pro His Gln Leu Arg Leu Arg Asn Val Ala Pro Ser Glu Val Ser Ser  
 2120                  2125                  2130

GAG GTG TCC ATA GAC ATT GGG ACC GAG ACT GAA GAC TCA GAA CTG ACT        6725  
 Glu Val Ser Ile Asp Ile Gly Thr Glu Thr Glu Asp Ser Glu Leu Thr  
 2135                  2140                  2145                  2150

GAG GCC GAC CTG CCG CCG GCA GCT GCA GCC CTC CAG GCT ATC GAG AAT        6773  
 Glu Ala Asp Leu Pro Pro Ala Ala Ala Leu Gln Ala Ile Glu Asn  
 2155                  2160                  2165

GCT GCG AGG ATT CTT GAG CCT CAT ATT GAT GTC ATC ATG GAG GAT TGC        6821  
 Ala Ala Arg Ile Leu Glu Pro His Ile Asp Val Ile Met Glu Asp Cys  
 2170                  2175                  2180

AGT ACA CCC TCT CTT TGT GGT AGT AGC CGA GAG ATG CCT GTG TGG GGA        6869  
 Ser Thr Pro Ser Leu Cys Gly Ser Ser Arg Glu Met Pro Val Trp Gly  
 2185                  2190                  2195

GAA GAC ATC CCC CGC ACT CCA TCG CCA GCA CTT ATC TCG GTT ACC GAG        6917  
 Glu Asp Ile Pro Arg Thr Pro Ser Pro Ala Leu Ile Ser Val Thr Glu  
 2200                  2205                  2210

AGC AGC TCA GAT GAG ARG ACC CCG TCG GTG TCC TCC TCG CAG GAG GAT        6965  
 Ser Ser Ser Asp Glu Lys Thr Pro Ser Val Ser Ser Gln Glu Asp  
 2215                  2220                  2225                  2230

ACC CCG TCC TCT GAC TCA TTC GAA GTC ATC CAA GAG TCT GAG ACA GCT        7013  
 Thr Pro Ser Ser Asp Ser Phe Glu Val Ile Gln Glu Ser Glu Thr Ala  
 2235                  2240                  2245

265

GAA GGA GAG GAA AGT GTC TTC AAC GTG GCT CTT TCC GTA CTA GAA CC Glu Gly Glu Glu Ser Val Phe Asn Val Ala Leu Ser Val Leu Glu Ala	7061
2250                    2255                    2260	
TTG TTT CCA CAG AGT GAT GCC ACT AGA AAG CTT ACC GTC AGG ATG AAT Leu Phe Pro Gln Ser Asp Ala Thr Arg Lys Leu Thr Val Arg Met Asn	7109
2265                    2270                    2275	
TGC TGC GTT GAG AAG AGC GTC ACG CGC TTC TTT TCT TTG GGG CTG ACG Cys Cys Val Glu Lys Ser Val Thr Arg Phe Phe Ser Leu Gly Leu Thr	7157
2280                    2285                    2290	
GTG GCT GAT GTG GCC AGT CTG TGT GAG ATG GAG ATC CAG AAC CAT ACA Val Ala Asp Val Ala Ser Leu Cys Glu Met Glu Ile Gln Asn His Thr	7205
2295                    2300                    2305                    2310	
GCC TAT TGT GAC AAG GTG CGC ACT CCG CTC GAA TTG CAA GTT GGG TGC Ala Tyr Cys Asp Lys Val Arg Thr Pro Leu Glu Leu Gln Val Gly Cys	7253
2315                    2320                    2325	
TTG GTG GGC AAT GAA CTT ACC TTT GAA TGT GAT AAG TGT GAG GCT AGG Leu Val Gly Asn Glu Leu Thr Phe Glu Cys Asp Lys Cys Glu Ala Arg	7301
2330                    2335                    2340	
CAA GAG ACT TTG GCC TCC TTC TCC TAT ATT TGG TCT GGG GTG CCA TTG Gln Glu Thr Leu Ala Ser Phe Ser Tyr Ile Trp Ser Gly Val Pro Leu	7349
2345                    2350                    2355	
ACT AGG GCC ACA CCG GCT AAA CCA CCT GTG GTG AGG CCG GTG GGG TCC Thr Arg Ala Thr Pro Ala Lys Pro Pro Val Val Arg Pro Val Gly Ser	7397
2360                    2365                    2370	
TTG TTG GTG GCT GAC ACC ACG AAA GTG TAT GTC ACA AAC CCG GAC AAT Leu Leu Val Ala Asp Thr Thr Lys Val Tyr Val Thr Asn Pro Asp Asn	7445
2375                    2380                    2385                    2390	
GTT GGG AGA AGA GTG GAC AAG GTG ACC TTC TGG CGC GCC CCC AGG GTC Val Gly Arg Arg Val Asp Lys Val Thr Phe Trp Arg Ala Pro Arg Val	7493
2395                    2400                    2405	
CAT GAC AAA TAT CTC GTG GAC TCC ATC GAG CGT GCC AGG AGG GCG GCT His Asp Lys Tyr Leu Val Asp S r Ile Glu Arg Ala Arg Arg Ala Ala	7541
2410                    2415                    2420	

266

CAA GCC TGC CAA AGC ATG GGT TAC ACT TAT GAG GAA GCA ATA AGG ACT Gln Ala Cys Gln Ser Met Gly Tyr Thr Tyr Glu Ile Ala Ile Arg Thr	2425	2430	2435	7589
GTT AGG CCA CAT GCT GCC ATG GGC TCG GGA TCT AAG GTG TCG GTC AAG Val Arg Pro His Ala Ala Met Gly Trp Gly Ser Lys Val Ser Val Lys	2440	2445	2450	7637
GAC TTG GCC ACC CCT GCG GGG AAG ATG GCC GTC CAC GAC CGA CTT CAG Asp Leu Ala Thr Pro Ala Gly Lys Met Ala Val His Asp Arg Leu Gln	2455	2460	2465	2470
GAG ATA CTT GAG GGG ACT CCG GTC CCT TTT ACT CTT ACT GTG AAA AAG Glu Ile Leu Glu Gly Thr Pro Val Pro Phe Thr Leu Thr Val Lys Lys	2475	2480	2485	7733
GAG GTG TTC TTC AAA GAC CGT AAG GAG GAG AAG GCC CCC CGC CTC ATT Glu Val Phe Phe Lys Asp Arg Lys Glu Glu Lys Ala Pro Arg Leu Ile	2490	2495	2500	7781
GTG TTC CCC CCC CTG GAC TTC CGG ATA GCT GAG AAG CTT ATC CTG GGA Val Phe Pro Pro Leu Asp Phe Arg Ile Ala Glu Lys Leu Ile Leu Gly	2505	2510	2515	7829
GAC CCG GGG CGG GTG GCC AAG GCG GTG TTG GGG GGG GCT TAC GCC TTC Asp Pro Gly Arg Val Ala Lys Ala Val Leu Gly Gly Ala Tyr Ala Phe	2520	2525	2530	7877
CAG TAC ACC CCA AAT CAG CGA GTT AAG GAG ATG CTC AAA CTG TGG GAG Gln Tyr Thr Pro Asn Gln Arg Val Lys Glu Met Leu Lys Leu Trp Glu	2535	2540	2545	2550
TCA AAG AAA ACA CCT TGC GCC ATC TGT GTG GAC GCC ACT TGC TTC GAC Ser Lys Lys Thr Pro Cys Ala Ile Cys Val Asp Ala Thr Cys Phe Asp	2555	2560	2565	7973
ACT AGC ATT ACT GAA GAG GAC GTG GCG CTG GAG ACA GAG CTG TAC GCT Ser Ser Ile Thr Glu Glu Asp Val Ala Leu Glu Thr Glu Leu Tyr Ala	2570	2575	2580	8021
CTG GCC TCT GAC CAT CCA GAG TGG GTG CGA GCT TTG GGG AAG TAC TAT Leu Ala Ser Asp His Pro Glu Trp Val Arg Ala Leu Gly Lys Tyr Tyr	2585	2590	2595	8069

267

GCC TCA GGA ACC ATG GTC ACC CCT GAG GGG GTT CCC GTA GGT GAG AGC Ala Ser Gly Thr Met Val Thr Pro Glu Gly Val Pro Val Gly Glu Arg	8117
2600                    2605                    2610	
TAT TGT AGA TCC TCA GGC GTT TTG ACT ACC AGC GCG AGT AAC TGC CTG Tyr Cys Arg Ser Ser Gly Val Leu Thr Thr Ser Ala Ser Asn Cys Leu	8165
2615                    2620                    2625                    2630	
ACC TGC TAC ATC AAG GTG AAA GCC GCT TGT GAG AGA GTG GGG CTG AAA Thr Cys Tyr Ile Lys Val Lys Ala Ala Cys Glu Arg Val Gly Leu Lys	8213
2635                    2640                    2645	
AAT GTC TCG CTT CTC ATA GCC GGC GAT GAC TGT TTG ATC ATA TGC GAA Asn Val Ser Leu Leu Ile Ala Gly Asp Asp Cys Leu Ile Ile Cys Glu	8261
2650                    2655                    2660	
CGG CCA GTG TGC GAC CCT TGT GAC GCC TTG GGC AGA GCC CTG GCG AGC Arg Pro Val Cys Asp Pro Cys Asp Ala Leu Gly Arg Ala Leu Ala Ser	8309
2665                    2670                    2675	
TAT GGG TAT GCT TGC GAG CCT TCG TAT CAT GCA TCA CTG GAC ACG GCC Tyr Gly Tyr Ala Cys Glu Pro Ser Tyr His Ala Ser Leu Asp Thr Ala	8357
2680                    2685                    2690	
CCC TTC TGC TCC ACT TGG CTC GCT GAG TGC AAC GCA GAT GGG AAA CGC Pro Phe Cys Ser Thr Trp Leu Ala Glu Cys Asn Ala Asp Gly Lys Arg	8405
2695                    2700                    2705                    2710	
CAT TTC TTC CTG ACC ACG GAC TTT CGG AGG CCG CTT GCT CGC ATG TCG His Phe Phe Leu Thr Thr Asp Phe Arg Arg Pro Leu Ala Arg Met Ser	8453
2715                    2720                    2725	
AGC GAG TAT AGT GAC CCA ATG GCT TCG GCC ATA GGT TAC ATC CTC CTG Ser Glu Tyr Ser Asp Pro Met Ala Ser Ala Ile Gly Tyr Ile Leu Leu	8501
2730                    2735                    2740	
TAT CCC TGG CAT CCC ATC ACA CGG TGG GTC ATC ATC CCT CAT GTG CTA Tyr Pro Trp His Pro Ile Thr Arg Trp Val Ile Ile Pro His Val Leu	8549
2745                    2750                    2755	
ACG TGC GCA TTC AGG GGT GGT ACA CCG TCT GAT CCG GTT TGG TGT Thr Cys Ala Phe Arg Gly Gly Thr Pro Ser Asp Pro Val Trp Cys	8597
2760                    2765                    2770	

268

CAG GTG CAT GGT AAC TAC TAC AAG TTT CCA CTG GAC AAA CTG CCT AAC Gln Val His Gly Asn Tyr Tyr Lys Phe Pro Leu Asp Lys L u Pro Asn	8645
2775                   2780                   2785                   2790	
 ATC ATC GTG GCC CTC CAC GGA CCA GCA GCG TTG AGG GTT ACC GCA GAC Ile Ile Val Ala Leu His Gly Pro Ala Ala Leu Arg Val Thr Ala Asp	8693
2795                   2800                   2805	
 ACA ACT AAG ACA AAA ATG GAA GCT GGG AAG GTG CTG AGT GAC CTC AAG Thr Thr Lys Thr Lys Met Glu Ala Gly Lys Val Leu Ser Asp Leu Lys	8741
2810                   2815                   2820	
 CTC CCT GGC CTA GCG GTC CAC CGA AAG AAG GCC GGA GCA CTG CGA ACA Leu Pro Gly Leu Ala Val His Arg Lys Lys Ala Gly Ala Leu Arg Thr	8789
2825                   2830                   2835	
 CGC ATG CTT CGG TCG CGC GGT TGG GCC GAG TTG GCG AGG GGC CTG TTG Arg Met Leu Arg Ser Arg Gly Trp Ala Glu Leu Ala Arg Gly Leu Leu	8837
2840                   2845                   2850	
 TGG CAT CCA GGC CTC CGG CTC CCT CCC CCT GAG ATT GCT GGT ATC CCG Trp His Pro Gly Leu Arg Leu Pro Pro Pro Glu Ile Ala Gly Ile Pro	8885
2855                   2860                   2865                   2870	
 GGG GGT TTC CCC CTC TCC CCC CCC TAC ATG GGG GTG GTG CAT CAA TTG Gly Gly Phe Pro Leu Ser Pro Pro Tyr Met Gly Val Val His Gln Leu	8933
2875                   2880                   2885	
 GAT TTT ACA AGC CAG AGG AGT CGC TGG CGG TGG CTG GGG TTC TTA GCC Asp Phe Thr Ser Gln Arg Ser Arg Trp Arg Trp Leu Gly Phe Leu Ala	8981
2890                   2895                   2900	
 CTG CTC ATC GTA GCC CTC TTC GGG TGAACTAAAT TCATCTGTTG CGGCAAGGTC Leu Leu Ile Val Ala Leu Phe Gly	9035
2905                   2910	
 CAGTGACTGA TCATCACTGG AGGAGGTTCC CGCCCTCCCC GCCCCAGGGG TCTCCCCGCT	9095
 GGGTAAAAA	9103

(2) INFORMATION FOR SEQ ID NO:157:

(i) SEQUENCE CHARACTERISTICS:

269

- (A) LENGTH: 2910 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:157:

Met Ser Leu Leu Thr Asn Arg Leu Ser Arg Arg Val Asp Lys Asp Gln  
1 5 10 15

Trp Gly Pro Gly Phe Met Gly Lys Asp Pro Lys Pro Cys Pro Ser Arg  
20 25 30

Arg Thr Gly Lys Cys Met Gly Pro Pro Ser Ser Ala Ala Ala Cys Ser  
35 40 45

Arg Gly Ser Pro Arg Ile Leu Arg Val Arg Ala Gly Gly Ile Ser Leu  
50 55 60

Pro Tyr Thr Ile Met Glu Ala Leu Leu Phe Leu Leu Gly Val Glu Ala  
65 70 75 80

Gly Ala Ile Leu Ala Pro Ala Thr His Ala Cys Arg Ala Asn Gly Gln  
85 90 95

Tyr Phe Leu Thr Asn Cys Cys Ala Pro Glu Asp Ile Gly Phe Cys Leu  
100 105 110

Glu Gly Gly Cys Leu Val Ala Leu Gly Cys Thr Val Cys Thr Asp Arg  
115 120 125

Cys Trp Pro Leu Tyr Gln Ala Gly Leu Ala Val Arg Pro Gly Lys Ser  
130 135 140

Ala Ala Gln Leu Val Gly Gln Leu Gly Gly Leu Tyr Gly Pro Leu Ser  
145 150 155 160

Val Ser Ala Tyr Val Ala Gly Ile Leu Gly Leu Gly Glu Val Tyr Ser  
165 170 175

Gly Val Leu Thr Val Gly Val Ala Leu Thr Arg Arg Val Tyr Pro Met  
180 185 190

270

Pro Asn Leu Thr Cys Ala Val Glu Cys Glu Leu Lys Trp Glu Ser Glu  
195 200 205

Phe Trp Arg Trp Thr Glu Gln Leu Ala Ser Asn Tyr Trp Ile Leu Glu  
210 215 220

Tyr Leu Trp Lys Val Pro Phe Asp Phe Trp Arg Gly Val Leu Ser Leu  
225 230 235 240

Thr Pro Leu Leu Val Cys Val Ala Ala Leu Leu Leu Glu Gln Arg  
245 250 255

Ile Val Met Val Phe Leu Leu Val Thr Met Ala Gly Met Ser Gln Gly  
260 265 270

Ala Pro Ala Ser Val Leu Gly Ser Arg Pro Phe Asp Tyr Gly Leu Thr  
275 280 285

Trp Gln Ser Cys Ser Cys Arg Ala Asn Gly Ser Arg Tyr Thr Thr Gly  
290 295 300

Glu Lys Val Trp Asp Arg Gly Asn Val Thr Leu Leu Cys Asp Cys Pro  
305 310 315 320

Asn Gly Pro Trp Val Trp Leu Pro Ala Phe Cys Gln Ala Ile Gly Trp  
325 330 335

Gly Asp Pro Ile Thr His Trp Ser His Gly Gln Asn Arg Trp Pro Leu  
340 345 350

Ser Cys Pro Gln Tyr Val Tyr Gly Ser Val Ser Val Thr Cys Val Trp  
355 360 365

Gly Ser Val Ser Trp Phe Ala Ser Thr Gly Gly Arg Asp Ser Lys Ile  
370 375 380

Asp Val Trp Ser Leu Val Pro Val Gly Ser Ala Ser Cys Thr Ile Ala  
385 390 395 400

Ala Leu Gly Ser Ser Asp Arg Asp Thr Val Val Glu Leu Ser Glu Trp  
405 410 415

Gly Val Pro Cys Ala Thr Cys Ile Leu Asp Arg Arg Pro Ala Ser Cys  
420 425 430

271

Gly Thr Cys Val Arg Asp Cys Trp Pr Glu Thr Gly Ser Val Arg Phe  
435 440 445

Pro Phe His Arg Cys Gly Ala Gly Pro Lys Leu Thr Lys Asp Leu Glu  
450 455 460

Ala Val Pro Phe Val Asn Arg Thr Thr Pro Phe Thr Ile Arg Gly Pro  
465 470 475 480

Leu Gly Asn Gln Gly Arg Gly Asn Pro Val Arg Ser Pro Leu Gly Phe  
485 490 495

Gly Ser Tyr Ala Met Thr Lys Ile Arg Asp Ser Leu His Leu Val Lys  
500 505 510

Cys Pro Thr Pro Ala Ile Glu Pro Pro Thr Gly Thr Phe Gly Phe Phe  
515 520 525

Pro Gly Val Pro Pro Leu Asn Asn Cys Leu Leu Leu Gly Thr Glu Val  
530 535 540

Ser Glu Ala Leu Gly Gly Ala Gly Leu Thr Gly Gly Phe Tyr Glu Pro  
545 550 555 560

Leu Val Arg Arg Arg Ser Glu Leu Met Gly Arg Arg Asn Pro Val Cys  
565 570 575

Pro Gly Phe Ala Trp Leu Ser Ser Gly Arg Pro Asp Gly Phe Ile His  
580 585 590

Val Gln Gly His Leu Gln Glu Val Asp Ala Gly Asn Phe Ile Pro Pro  
595 600 605

Pro Arg Trp Leu Leu Leu Asp Phe Val Phe Val Leu Leu Tyr Leu Met  
610 615 620

Lys Leu Ala Glu Ala Arg Leu Val Pro Leu Ile Leu Leu Leu Trp  
625 630 635 640

Trp Trp Val Asn Gln Leu Ala Val Leu Gly Leu Pro Ala Val Asp Ala  
645 650 655

Ala Val Ala Gly Glu Val Phe Ala Gly Pro Ala Leu Ser Trp Cys Leu  
660 665 670

272

Gly Leu Pro Thr Val Ser Met Ile Leu Gly Leu Ala Asn Leu Val Leu  
675 680 685

Tyr Phe Arg Trp Met Gly Pro Gln Arg Leu Met Phe Leu Val Leu Trp  
690 695 700

Lys Leu Ala Arg Gly Ala Phe Pro Leu Ala Leu Leu Met Gly Ile Ser  
705 710 715 720

Ala Thr Arg Gly Arg Thr Ser Val Leu Gly Ala Glu Phe Cys Phe Asp  
725 730 735

Val Thr Phe Glu Val Asp Thr Ser Val Leu Gly Trp Val Val Ala Ser  
740 745 750

Val Val Ala Trp Ala Ile Ala Leu Leu Ser Ser Met Ser Ala Gly Gly  
755 760 765

Trp Arg His Lys Ala Val Ile Tyr Arg Thr Trp Cys Lys Gly Tyr Gln  
770 775 780

Ala Ile Arg Gln Arg Val Val Arg Ser Pro Leu Gly Glu Gly Arg Pro  
785 790 795 800

Thr Lys Pro Leu Thr Phe Ala Trp Cys Leu Ala Ser Tyr Ile Trp Pro  
805 810 815

Asp Ala Val Met Met Val Val Val Ala Leu Val Leu Leu Phe Gly Leu  
820 825 830

Phe Asp Ala Leu Asp Trp Ala Leu Glu Glu Leu Leu Val Ser Arg Pro  
835 840 845

Ser Leu Arg Arg Leu Ala Arg Val Val Glu Cys Cys Val Met Ala Gly  
850 855 860

Glu Lys Ala Thr Thr Val Arg Leu Val Ser Lys Met Cys Ala Arg Gly  
865 870 875 880

Ala Tyr Leu Phe Asp His Met Gly Ser Phe Ser Arg Ala Val Lys Glu  
885 890 895

Arg Leu Leu Glu Trp Asp Ala Ala Leu Glu Pro Leu Ser Ph Thr Arg  
900 905 910

273

Thr Asp Cys Arg Ile Ile Arg Asp Ala Ala Arg Thr Leu Ala Cys Gly  
915                    920                    925

Gln Cys Val Met Gly Leu Pro Val Val Ala Arg Arg Gly Asp Glu Val  
930                    935                    940

Leu Ile Gly Val Phe Gln Asp Val Asn His Leu Pro Pro Gly Phe Val  
945                    950                    955                    960

Pro Thr Ala Pro Val Val Ile Arg Arg Cys Gly Lys Gly Phe Leu Gly  
965                    970                    975

Val Thr Lys Ala Ala Leu Thr Gly Arg Asp Pro Asp Leu His Pro Gly  
980                    985                    990

Asn Val Met Val Leu Gly Thr Ala Thr Ser Arg Ser Met Gly Thr Cys  
995                    1000                    1005

Leu Asn Gly Leu Leu Phe Thr Thr Phe His Gly Ala Ser Ser Arg Thr  
1010                    1015                    1020

Ile Ala Thr Pro Val Gly Ala Leu Asn Pro Arg Trp Trp Ser Ala Ser  
1025                    1030                    1035                    1040

Asp Asp Val Thr Val Tyr Pro Leu Pro Asp Gly Ala Thr Ser Leu Thr  
1045                    1050                    1055

Pro Cys Thr Cys Gln Ala Glu Ser Cys Trp Val Ile Arg Ser Asp Gly  
1060                    1065                    1070

Ala Leu Cys His Gly Leu Ser Lys Gly Asp Lys Val Glu Leu Asp Val  
1075                    1080                    1085

Ala Met Glu Val Ser Asp Phe Arg Gly Ser Ser Gly Ser Pro Val Leu  
1090                    1095                    1100

Cys Asp Glu Gly His Ala Val Gly Met Leu Val Ser Val Leu His Ser  
1105                    1110                    1115                    1120

Gly Gly Arg Val Thr Ala Ala Arg Phe Thr Arg Pro Trp Thr Gln Val  
1125                    1130                    1135

Pro Thr Asp Ala Lys Thr Thr Glu Pro Pro Pro Val Pro Ala Lys  
1140                    1145                    1150

274

Gly Val Phe Lys Glu Ala Pro Leu Phe Met Pro Thr Gly Ala Gly Lys  
1155 1160 1165

Ser Thr Arg Val Pro Leu Glu Tyr Gly Asn Met Gly His Lys Val Leu  
1170 1175 1180

Ile Leu Asn Pro Ser Val Ala Thr Val Arg Ala Met Gly Pro Tyr Met  
1185 1190 1195 1200

Glu Arg Leu Ala Gly Lys His Pro Ser Ile Tyr Cys Gly His Asp Thr  
1205 1210 1215

Thr Ala Phe Thr Arg Ile Thr Asp Ser Pro Leu Thr Tyr Ser Thr Tyr  
1220 1225 1230

Gly Arg Phe Leu Ala Asn Pro Arg Gln Met Leu Arg Gly Val Ser Val  
1235 1240 1245

Val Ile Cys Asp Glu Cys His Ser His Asp Ser Thr Val Leu Leu Gly  
1250 1255 1260

Ile Gly Arg Val Arg Glu Leu Ala Arg Glu Cys Gly Val Gln Leu Val  
1265 1270 1275 1280

Leu Tyr Ala Thr Ala Thr Pro Pro Gly Ser Pro Met Thr Gln His Pro  
1285 1290 1295

Ser Ile Ile Glu Thr Lys Leu Asp Val Gly Glu Ile Pro Phe Tyr Gly  
1300 1305 1310

His Gly Ile Pro Leu Glu Arg Met Arg Thr Gly Arg His Leu Val Phe  
1315 1320 1325

Cys Tyr Ser Lys Ala Glu Cys Glu Arg Leu Ala Gly Gln Phe Ser Ala  
1330 1335 1340

Arg Gly Val Asn Ala Ile Ala Tyr Tyr Arg Gly Lys Asp Ser Ser Ile  
1345 1350 1355 1360

Ile Lys Asp Gly Asp Leu Val Val Cys Ala Thr Asp Ala Leu Ser Thr  
1365 1370 1375

Gly Tyr Thr Gly Asn Phe Asp Ser Val Thr Asp Cys Gly Leu Val Val  
1380 1385 1390

275

Glu Glu Val Val Glu Val Thr Leu Asp Pro Thr Ile Thr Ile Ser Leu  
1395 1400 1405

Arg Thr Val Pro Ala Ser Ala Glu Leu Ser Met Gln Arg Arg Gly Arg  
1410 1415 1420

Thr Gly Arg Gly Arg Ser Gly Arg Tyr Tyr Tyr Ala Gly Val Gly Lys  
1425 1430 1435 1440

Ala Pro Ala Gly Val Val Arg Ser Gly Pro Val Trp Ser Ala Val Glu  
1445 1450 1455

Ala Gly Val Thr Trp Tyr Gly Met Glu Pro Asp Leu Thr Ala Asn Leu  
1460 1465 1470

Leu Arg Leu Tyr Asp Asp Cys Pro Tyr Thr Ala Ala Val Ala Ala Asp  
1475 1480 1485

Ile Gly Glu Ala Ala Val Phe Phe Ser Gly Leu Ala Pro Leu Arg Met  
1490 1495 1500

His Pro Asp Val Ser Trp Ala Lys Val Arg Gly Val Asn Trp Pro Leu  
1505 1510 1515 1520

Leu Val Gly Val Gln Arg Thr Met Cys Arg Glu Thr Leu Ser Pro Gly  
1525 1530 1535

Pro Ser Asp Asp Pro Gln Trp Ala Gly Leu Lys Gly Pro Asn Pro Val  
1540 1545 1550

Pro Leu Leu Leu Arg Trp Gly Asn Asp Leu Pro Ser Lys Val Ala Gly  
1555 1560 1565

His His Ile Val Asp Asp Leu Val Arg Arg Leu Gly Val Ala Glu Gly  
1570 1575 1580

Tyr Val Arg Cys Asp Ala Gly Pro Ile Leu Met Val Gly Leu Ala Ile  
1585 1590 1595 1600

Ala Gly Gly Met Ile Tyr Ala Ser Tyr Thr Gly Ser Leu Val Val Val  
1605 1610 1615

Thr Asp Trp Asp Val Lys Gly Gly Ser Pro Leu Tyr Arg His Gly  
1620 1625 1630

276

Asp Gln Ala Thr Pro Gln Pro Val Val Gln Val Pro Pro Val Asp His  
1635                    1640                    1645

Arg Pro Gly Gly Glu Ser Ala Pro Ser Asp Ala Lys Thr Val Thr Asp  
1650                    1655                    1660

Ala Val Ala Ala Ile Gln Val Asp Cys Asp Trp Ser Val Met Thr Leu  
1665                    1670                    1675                    1680

Ser Ile Gly Glu Val Leu Ser Leu Ala Gln Ala Lys Thr Ala Glu Ala  
1685                    1690                    1695

Tyr Thr Ala Thr Ala Lys Trp Leu Ala Gly Cys Tyr Thr Gly Thr Arg  
1700                    1705                    1710

Ala Val Pro Thr Val Ser Ile Val Asp Lys Leu Phe Ala Gly Gly Trp  
1715                    1720                    1725

Ala Ala Val Val Gly His Cys His Ser Val Ile Ala Ala Ala Val Ala  
1730                    1735                    1740

Ala Tyr Gly Ala Ser Arg Ser Pro Pro Leu Ala Ala Ala Ala Ser Tyr  
1745                    1750                    1755                    1760

Leu Met Gly Leu Gly Val Gly Gly Asn Ala Gln Thr Arg Leu Ala Ser  
1765                    1770                    1775

Ala Leu Leu Leu Gly Ala Ala Gly Thr Ala Leu Gly Thr Pro Val Val  
1780                    1785                    1790

Gly Leu Thr Met Ala Gly Ala Phe Met Gly Gly Ala Ser Val Ser Pro  
1795                    1800                    1805

Ser Leu Val Thr Ile Leu Leu Gly Ala Val Gly Gly Trp Glu Gly Val  
1810                    1815                    1820

Val Asn Ala Ala Ser Leu Val Phe Asp Phe Met Ala Gly Lys Leu Ser  
1825                    1830                    1835                    1840

Ser Glu Asp Leu Trp Tyr Ala Ile Pro Val Leu Thr Ser Pro Gly Ala  
1845                    1850                    1855

ly Leu Ala Gly Ile Ala Leu Gly Leu Val Leu Tyr Ser Ala Asn Asn  
1860                    1865                    1870

277

Ser Gly Thr Thr Trp Leu Asn Arg Leu Leu Thr Thr Leu Pro Arg  
1875 1880 1885

Ser Ser Cys Ile Pro Asp Ser Tyr Phe Gln Gln Ala Asp Tyr Cys Asp  
1890 1895 1900

Lys Val Ser Ala Val Leu Arg Arg Leu Ser Leu Thr Arg Thr Val Val  
1905 1910 1915 1920

Ala Leu Val Asn Arg Glu Pro Lys Val Asp Glu Val Gln Val Gly Tyr  
1925 1930 1935

Val Trp Asp Leu Trp Glu Trp Ile Met Arg Gln Val Arg Met Val Met  
1940 1945 1950

Ala Arg Leu Arg Ala Leu Cys Pro Val Val Ser Leu Pro Leu Trp His  
1955 1960 1965

Cys Gly Glu Gly Trp Ser Gly Glu Trp Leu Leu Asp Gly His Val Glu  
1970 1975 1980

Ser Arg Cys Leu Cys Gly Cys Val Ile Thr Gly Asp Val Phe Asn Gly  
1985 1990 1995 2000

Gln Leu Lys Glu Pro Val Tyr Ser Thr Lys Leu Cys Arg His Tyr Trp  
2005 2010 2015

Met Gly Thr Val Pro Val Asn Met Leu Gly Tyr Gly Glu Thr Ser Pro  
2020 2025 2030

Leu Leu Ala Ser Asp Thr Pro Lys Val Val Pro Phe Gly Thr Ser Gly  
2035 2040 2045

Trp Ala Glu Val Val Val Thr Pro Thr His Val Val Ile Arg Arg Thr  
2050 2055 2060

Ser Pro Tyr Glu Leu Leu Arg Gln Gln Ile Leu Ser Ala Ala Val Ala  
2065 2070 2075 2080

Glu Pro Tyr Tyr Val Asp Gly Ile Pro Val Ser Trp Asp Ala Asp Ala  
2085 2090 2095

Arg Ala Pro Ala Met Val Tyr Gly Pro Gly Gln Ser Val Thr Ile Asp  
2100 2105 2110

278

Gly Glu Arg Tyr Thr Leu Pro His Gln Leu Arg Leu Arg Asn Val Ala  
2115 2120 2125

Pro Ser Glu Val Ser Ser Glu Val Ser Ile Asp Ile Gly Thr Glu Thr  
2130 2135 2140

Glu Asp Ser Glu Leu Thr Glu Ala Asp Leu Pro Pro Ala Ala Ala  
2145 2150 2155 2160

Leu Gln Ala Ile Glu Asn Ala Ala Arg Ile Leu Glu Pro His Ile Asp  
2165 2170 2175

Val Ile Met Glu Asp Cys Ser Thr Pro Ser Leu Cys Gly Ser Ser Arg  
2180 2185 2190

Glu Met Pro Val Trp Gly Glu Asp Ile Pro Arg Thr Pro Ser Pro Ala  
2195 2200 2205

Leu Ile Ser Val Thr Glu Ser Ser Ser Asp Glu Lys Thr Pro Ser Val  
2210 2215 2220

Ser Ser Ser Gln Glu Asp Thr Pro Ser Ser Asp Ser Phe Glu Val Ile  
2225 2230 2235 2240

Gln Glu Ser Glu Thr Ala Glu Gly Glu Ser Val Phe Asn Val Ala  
2245 2250 2255

Leu Ser Val Leu Glu Ala Leu Phe Pro Gln Ser Asp Ala Thr Arg Lys  
2260 2265 2270

Leu Thr Val Arg Met Asn Cys Cys Val Glu Lys Ser Val Thr Arg Phe  
2275 2280 2285

Phe Ser Leu Gly Leu Thr Val Ala Asp Val Ala Ser Leu Cys Glu Met  
2290 2295 2300

Glu Ile Gln Asn His Thr Ala Tyr Cys Asp Lys Val Arg Thr Pro Leu  
2305 2310 2315 2320

Glu Leu Gln Val Gly Cys Leu Val Gly Asn Glu Leu Thr Phe Glu Cys  
2325 2330 2335

Asp Lys Cys Glu Ala Arg Gln Glu Thr Leu Ala Ser Phe Ser Tyr Ile  
2340 2345 2350

279

Trp Ser Gly Val Pro Leu Thr Arg Ala Thr Pro Ala Lys Pro Pro Val  
2355 2360 2365

Val Arg Pro Val Gly Ser Leu Leu Val Ala Asp Thr Thr Lys Val Tyr  
2370 2375 2380

Val Thr Asn Pro Asp Asn Val Gly Arg Arg Val Asp Lys Val Thr Phe  
2385 2390 2395 2400

Trp Arg Ala Pro Arg Val His Asp Lys Tyr Leu Val Asp Ser Ile Glu  
2405 2410 2415

Arg Ala Arg Arg Ala Ala Gln Ala Cys Gln Ser Met Gly Tyr Thr Tyr  
2420 2425 2430

Glu Glu Ala Ile Arg Thr Val Arg Pro His Ala Ala Met Gly Trp Gly  
2435 2440 2445

Ser Lys Val Ser Val Lys Asp Leu Ala Thr Pro Ala Gly Lys Met Ala  
2450 2455 2460

Val His Asp Arg Leu Gln Glu Ile Leu Glu Gly Thr Pro Val Pro Phe  
2465 2470 2475 2480

Thr Leu Thr Val Lys Lys Glu Val Phe Phe Lys Asp Arg Lys Glu Glu  
2485 2490 2495

Lys Ala Pro Arg Leu Ile Val Phe Pro Pro Leu Asp Phe Arg Ile Ala  
2500 2505 2510

Glu Lys Leu Ile Leu Gly Asp Pro Gly Arg Val Ala Lys Ala Val Leu  
2515 2520 2525

Gly Gly Ala Tyr Ala Phe Gln Tyr Thr Pro Asn Gln Arg Val Lys Glu  
2530 2535 2540

Met Leu Lys Leu Trp Glu Ser Lys Lys Thr Pro Cys Ala Ile Cys Val  
2545 2550 2555 2560

Asp Ala Thr Cys Phe Asp Ser Ser Ile Thr Glu Glu Asp Val Ala Leu  
2565 2570 2575

Glu Thr Glu Leu Tyr Ala Leu Ala Ser Asp His Pro Glu Trp Val Arg  
2580 2585 2590

280

Ala Leu Gly Lys Tyr Tyr Ala Ser Gly Thr Met Val Thr Pro Glu Gly  
2595 2600 2605

Val Pro Val Gly Glu Arg Tyr Cys Arg Ser Ser Gly Val Leu Thr Thr  
2610 2615 2620

Ser Ala Ser Asn Cys Leu Thr Cys Tyr Ile Lys Val Lys Ala Ala Cys  
2625 2630 2635 2640

Glu Arg Val Gly Leu Lys Asn Val Ser Leu Leu Ile Ala Gly Asp Asp  
2645 2650 2655

Cys Leu Ile Ile Cys Glu Arg Pro Val Cys Asp Pro Cys Asp Ala Leu  
2660 2665 2670

Gly Arg Ala Leu Ala Ser Tyr Gly Tyr Ala Cys Glu Pro Ser Tyr His  
2675 2680 2685

Ala Ser Leu Asp Thr Ala Pro Phe Cys Ser Thr Trp Leu Ala Glu Cys  
2690 2695 2700

Asn Ala Asp Gly Lys Arg His Phe Phe Leu Thr Thr Asp Phe Arg Arg  
2705 2710 2715 2720

Pro Leu Ala Arg Met Ser Ser Glu Tyr Ser Asp Pro Met Ala Ser Ala  
2725 2730 2735

Ile Gly Tyr Ile Leu Leu Tyr Pro Trp His Pro Ile Thr Arg Trp Val  
2740 2745 2750

Ile Ile Pro His Val Leu Thr Cys Ala Phe Arg Gly Gly Thr Pro  
2755 2760 2765

Ser Asp Pro Val Trp Cys Gln Val His Gly Asn Tyr Tyr Lys Phe Pro  
2770 2775 2780

Leu Asp Lys Leu Pro Asn Ile Ile Val Ala Leu His Gly Pro Ala Ala  
2785 2790 2795 2800

Leu Arg Val Thr Ala Asp Thr Thr Lys Thr Lys Met Glu Ala Gly Lys  
2805 2810 2815

Val Leu Ser Asp Leu Lys Leu Pro Gly Leu Ala Val His Arg Lys Lys  
2820 2825 2830

281

Ala Gly Ala Leu Arg Thr Arg Met Leu Arg Ser Arg Gly Trp Ala Glu  
2835                   2840                   2845

Leu Ala Arg Gly Leu Leu Trp His Pro Gly Leu Arg Leu Pro Pro Pro  
2850                   2855                   2860

Glu Ile Ala Gly Ile Pro Gly Gly Phe Pro Leu Ser Pro Pro Tyr Met  
2865                   2870                   2875                   2880

Gly Val Val His Gln Leu Asp Phe Thr Ser Gln Arg Ser Arg Trp Arg  
2885                   2890                   2895

Trp Leu Gly Phe Leu Ala Leu Leu Ile Val Ala Leu Phe Gly  
2900                   2905                   2910

**IT IS CLAIMED:**

1. A purified polypeptide antigen encoded by the reverse-frame of a virus having an RNA genome, where said  
5 polypeptide antigen is specifically immunoreactive with serum infected with said RNA virus.

2. A polypeptide antigen of claim 1, where said virus is a single, positive strand RNA virus.

10

3. A polypeptide antigen of claim 2, where said virus is Hepatitis G Virus (HGV) or Hepatitis C Virus (HCV).

15

4. A polypeptide antigen of claim 3, where said virus is HGV and said polypeptide antigen or a polypeptide antigen containing fragment is encoded by the sequence presented as SEQ ID NO:19 or SEQ ID NO:28.

20

5. A polypeptide antigen of claim 3, where said virus is HCV and said polypeptide antigen or a polypeptide antigen containing fragment is derived from a sequence selected from the group consisting of SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, SEQ ID NO:144, SEQ  
25 ID NO:145 and SEQ ID NO:146.

30

6. A method of detecting serum infected with a virus having an RNA genome, comprising reacting serum with a substantially isolated polypeptide antigen of claim 1, and  
examining the polypeptide antigen for the presence of bound antibody.

35

7. A method of claim 6, wherein the polypeptide antigen is attached to a solid support, said reacting includes reacting the serum with the support, and subsequently reacting the support with a reporter-

labelled anti-human antibody, and said examining includes detecting the presence of reporter-labelled antibody on the solid support.

5        8. A monoclonal antibody specifically immunoreactive with a polypeptide antigen of claim 1.

10      9. A substantially isolated preparation of polyclonal antibodies specifically immunoreactive with a polypeptide antigen of claim 1.

15      10. A preparation of polyclonal antibodies of claim 9, where said polyclonal antibodies are prepared by affinity.

15      11. A method of identifying a polypeptide antigen that is specifically immunoreactive with antibodies against a selected virus having an RNA genome, comprising determining a first polynucleotide sequence corresponding to coding sequences for identifiable viral proteins for the selected virus, generating a second polynucleotide sequence complementary to the first polynucleotide encoding said identifiable viral proteins,

20      25      examining the said second polynucleotide for the presence of an open reading frame (ORF), identifying a polypeptide antigen encoded by said ORF that is specifically immunoreactive with antibodies against said virus.

30      12. A method of claim 11, where said first polynucleotide is the genomic strand of a single, positive strand RNA virus that encodes a polyprotein.

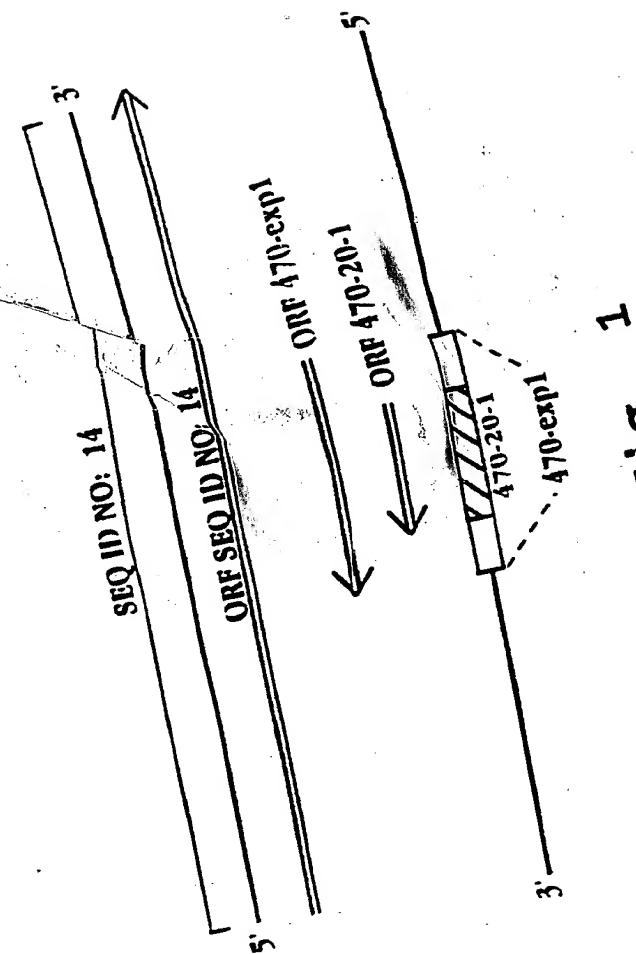
35      13. A method of claim 11, where said identifying includes producing said polypeptide antigen and screening

284

**said polypeptide antigen against sera infected with said virus.**

WO 95/32292

1/14



2/14

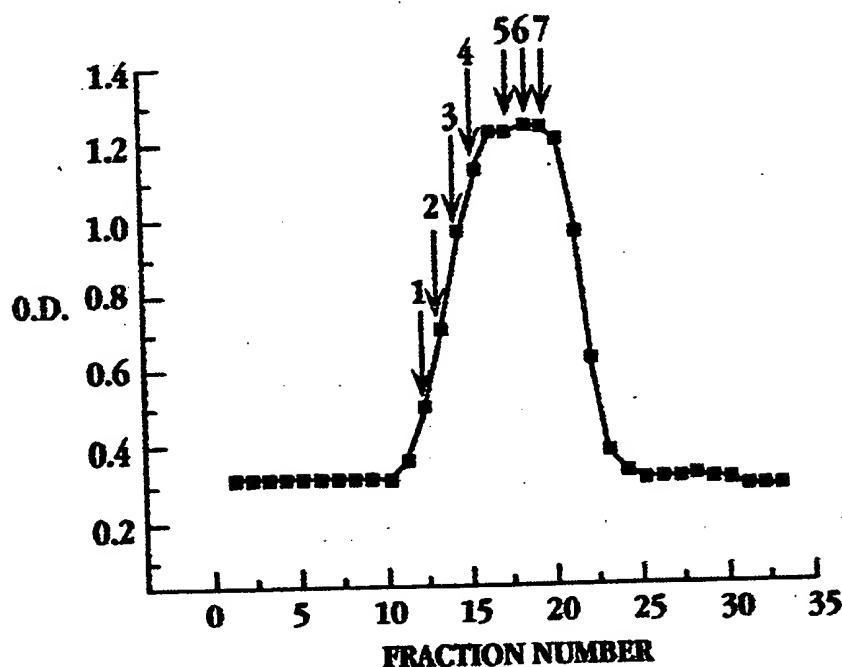


Fig. 2

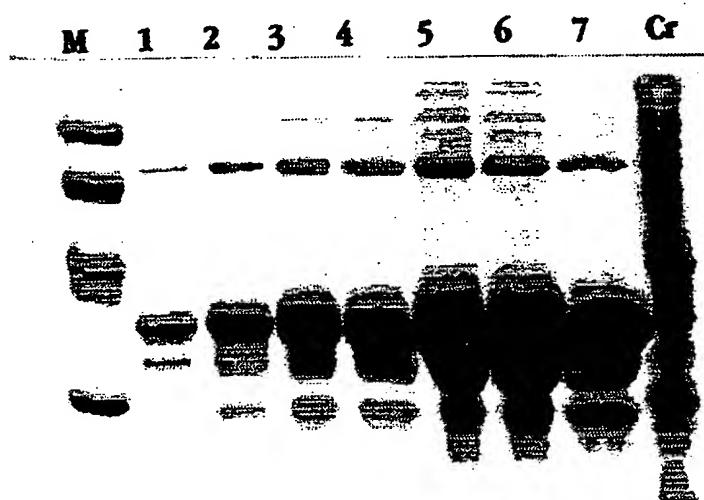
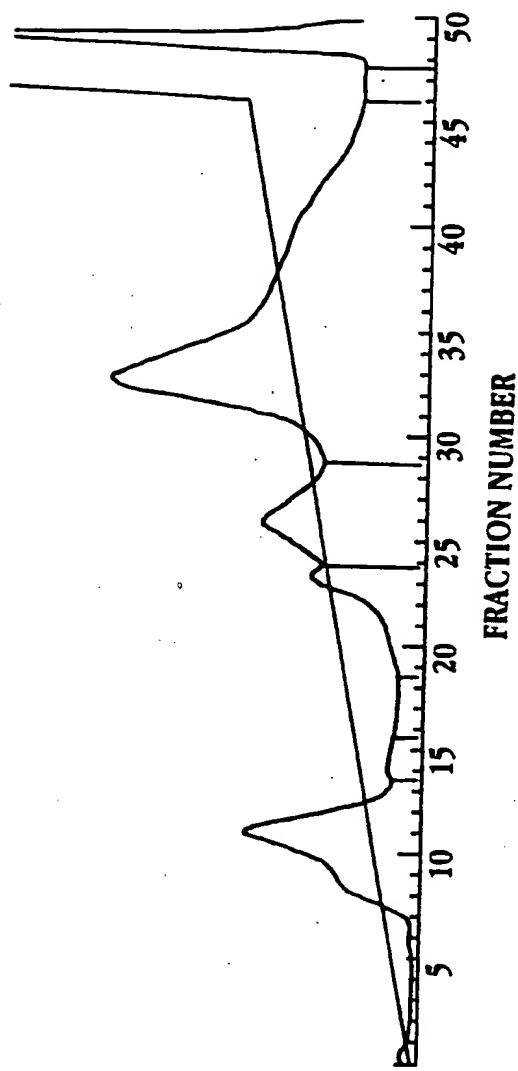


Fig. 3

3/14

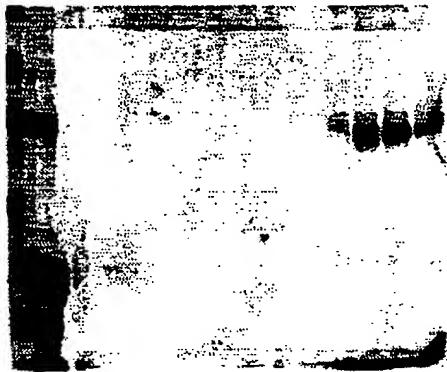


**Fig. 4A**

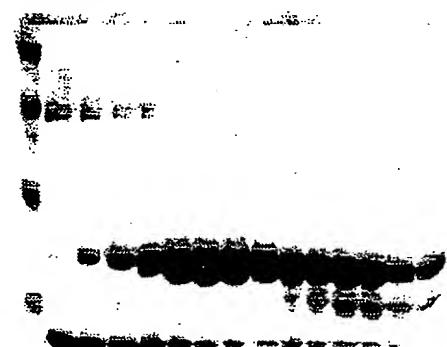
4/14

**FRACTION**

Crude 9 11 13 23 26 28  
8 10 12 14 24 27 29

**Fig. 4B****FRACTION**

30 32 34 36 38 40 50  
M 31 33 35 37 39 41 51

**POOL****Fig. 4C**

5/14

		10	20	30
HGV		* * * * *		
HOCV	ykwvkqkpvvipyegktplfqifdkvk 3490 3500	kewdqfqnpavsfDtkawDdtqvttndleik 3510 3520 3530 3540		
		40	50	60
HGV		* * * * *		
HOCV	el--yalasdhpewvrarpgkyyasgtnvtpegyvgerycrssgvlt 3550 3560	trsaNs-1tcyik digkyyfkkkwhkfiddtltmsevpvitadgevyirkqqRgsqqpdtsagNsmlnvltm 3570 3580 3590 3600		
		70	80	
HGV		* * * * *		
HOCV	90 100 110 120 130 140	vkaacervgl-----knvslliaGDDcllicerpvcdpsdalgralasygyacepsyha 3610 3620 3630 3640 3650 3660		
		90	100	110
HGV		* * * * *		
HOCV	vyafceatgypyksfdrvakihvcdDgfilteralgekfaskgyqillyeagkpqkiteq 3610 3620 3630 3640 3650 3660			

Fig. 5A

6/14

	10	20	30
HGV	lweskktptcaicvDatcfDssiteedvalet		
HCV	vvst1pqvvvmqssyfqysppgrvelfvntwkskknppmgfsyDtrcfDstytendirvee	2600 2610 2620 2630 2640 2650	
	40	50	60
HGV	elya---lasdhpevwra-pgkyyasgtmvtpiegvpvgerycrsBgvlttsasNcltcyl		
HCV	siYgccdiapearqaiksltterlyiggpltnskgqncgyrrRasBgvltscgNtltcyl	2660 2670 2680 2690 2700 2710	
	70	80	
HGV	kvkaacervglknvslliaGDDcliicerpvcdpsdalgralasygyacepsyhasldta		
HCV	kasaacraaklqdctmlvnGDDlvicesagtqedaaaslrvfteamtryssappgdpqpe	2720 2730 2740 2750 2760 2770	
	90	100	110
	***		
	120	130	140

Fig. 5B

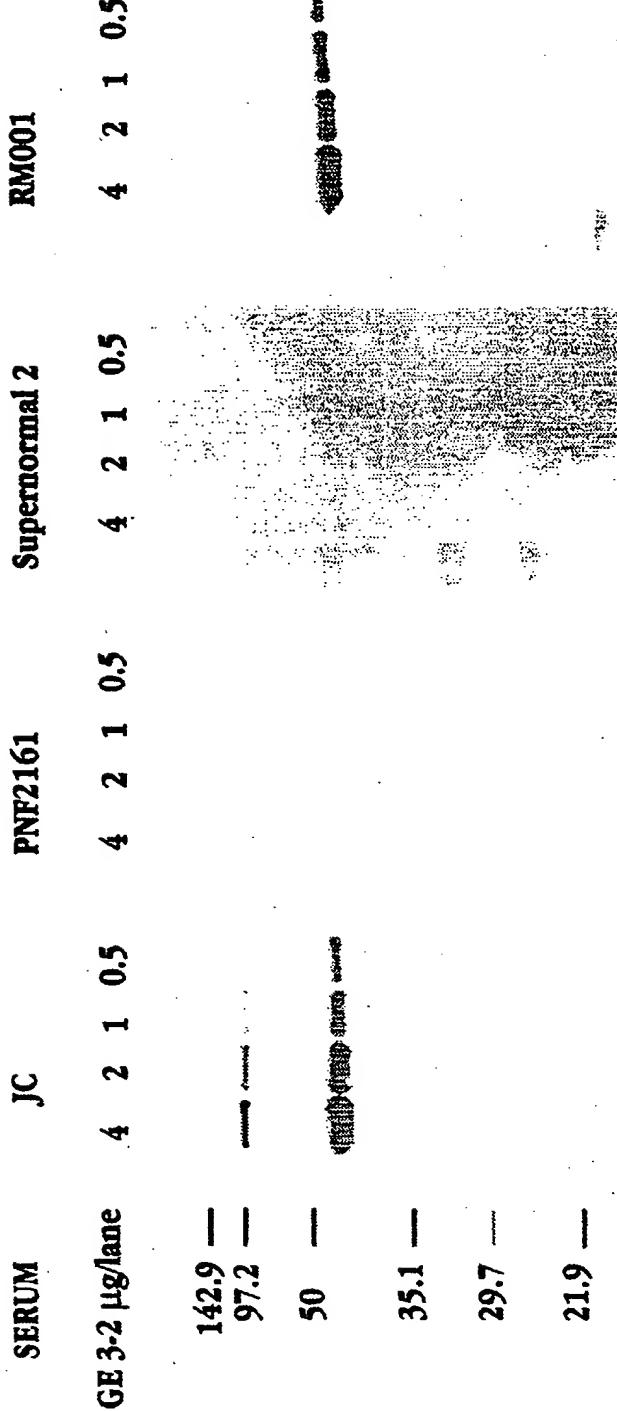
7/14

Thrombin cleavage

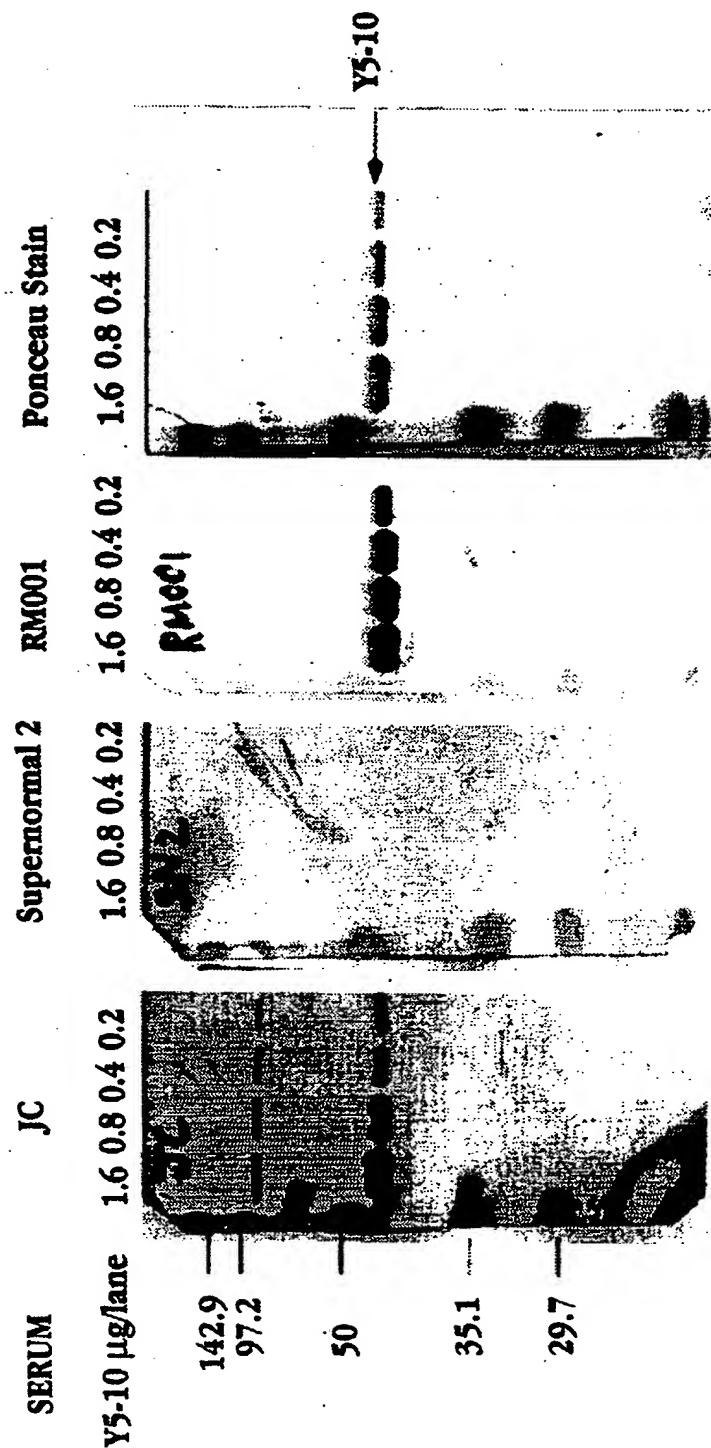
sj26	K S D L V P R G S M V S W D A D A R A P	GE3-2				
1	11	21	31	41	51	
CAAAATCGGATCTGGTCCGCGTGGTTCATGGTCTCATGGACGGACGCGCTCGTGC						C^CATGG (NcoI)
61	A M V Y G P G Q S V T I D G E R Y T L P	61	71 81 91 101 111			
CCGCGATGGTCTATGGCCCTGGGCAAAGTGTACCATTGACGGGAGCGCTACACCTTGC						^Base mutated to remove NcoI site AGC^GCT (Eco47III)
121	H Q L R L R N V A P S E V S S E V S I D	121	131 141 151 161 171			
CTCATCAACTGAGGCTCAGGAATGTGGCACCCCTCTGAGGTTTCATCCGAGGTGTCCATTG						
181	I G T E T E D S E L T E A D L P P A A A	181	191 201 211 221 231			
ACATTGGGACGGAGACTGAAGACTCAGAACTGACTGAGGCCGATCTGCCGCCGGCGCTG						CTGAAG (Eco57I_16/14->) GCC^GGC (NaeI)
CTTCAG (<-14/16_Eco57I)						
241	A L Q A I E N A A R I L E P H I D V I M	241	251 261 271 281 291			
CTGCTCTCCAAGCGATCGAGAATGCTGCGAGGATTCTTGAACCGCACATTGATGTCATCA						CGAT^CG (PvuI)
GAATGCN^ (BsmI)						
301	E D C S T P S L C G S S R E M P V W G E	301	311 321 331 341 351			
TGGAGGACTGCAGTACACCCCTCTTTGTGGTAGTAGCCGAGAGATGCCTGTATGGGGAG						CTGCA^G (PstI)
END-GE3-2>						
361	D I P R T P S P A L I G S H H H H H H Z	361	371 381 391 401 411			
AAGACATCCCCCGTACTCCATGCCAGCACTTATCGGATCCCACCATCACCATCACCATT						G^GATCC (BamHI)
pGEX						
421	N S S Z L T D D L P	421	431 441 451			
AGAATTCATCGTGAAGTGACTGACTGACGATCTACCT						G^AATTC (EcoR1)

Fig. 6

8/14

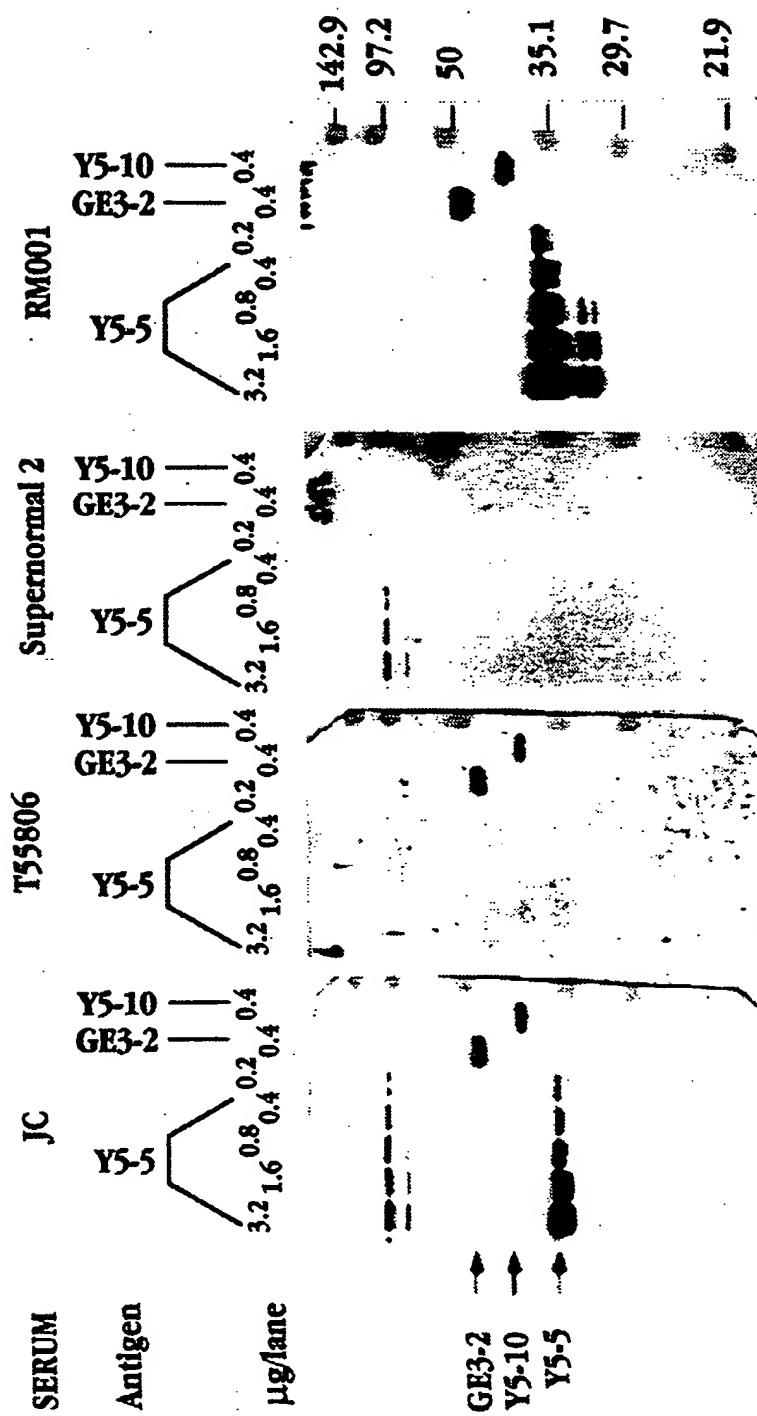
**Fig. 7A Fig. 7B Fig. 7C Fig. 7D**

9/14



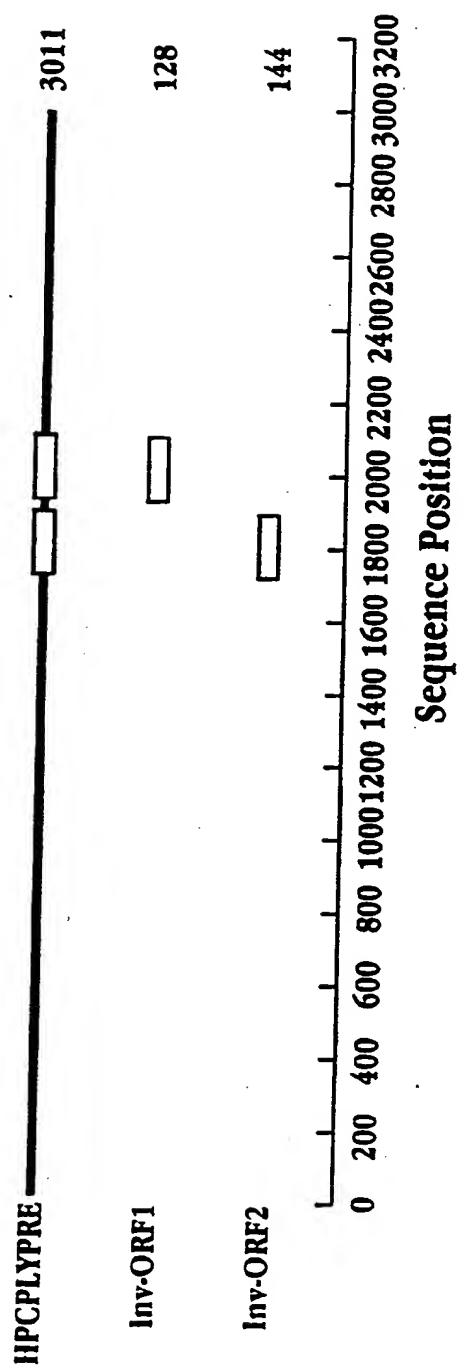
**Fig. 8A Fig. 8B Fig. 8C Fig. 8D**

10/14



**Fig. 9A Fig. 9B Fig. 9C Fig. 9D**

11/14



**Fig. 10**

12/14

Convene

Fig. 11A.

13/14

卷之三

P ALIELRSAMAQAIT	P ALIELRSAMAQATT												
k3 - 8 - 3	k3 - 8 - 5	k3 - 10 - 1d	k3 - 8 - 4	k3 - 8 - 7	k3 - 14 - 3	k3 - 14 - 6	k3 - 17 - 1	k3 - 14 - 2	k3 - 14 - 5	k3 - 11 - 1	k3 - 8 - 6		

P-ALIBELPSAMPAQATTLATTOPNTBVSTSNSVASKONSTPSTEVRPRVAEIPIKRASGKAPRA  
T I A

(con't) k3-8-6 123 SFHNTNIRRNGPNHLKYSTRREAKPNILITGSPRHQDRAGPAESTPAAASTAGSPNLZmr1

SFHNTRNRWOPNHLKYSTRFAKPNIILTGSPRHQDRAGPAETSPAAAATAGSPNL  
consensus

14/14

k3 - 8 - 3	QHSTNRASASISRECF 
k3 - 8 - 5	QHSTNRASASArsldan 
k3 - 10 - 1d	QHSTNRASASISRCEF 
k3 - 8 - 4	QHSTNRASASISRCEF 
k3 - 8 - 7	QHSTNRASASISRCEF 
k3 - 14 - 3	QHSTNRASASISRCEF 
k3 - 14 - 6	QHSTNRASASISRCEF 
k3 - 17 - 1	QHSTNRASASISRCEF 
k3 - 14 - 2	QHSTNRASASISRCEF 
k3 - 14 - 5	QHSTNRASASISRCEF 

consensus

TNRASAS

Fig. 11C